UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION This Document Relates To: News Cases 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	25-md-3143 (SHS) (OTW)

NEWS PLAINTIFFS' OMNIBUS REPLY MEMORANDUM OF LAW IN SUPPORT OF NEWS PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT (DKT. 1696)

REDACTED

TABLE OF CONTENTS

I. Introduction 1
II. Argument 3
A. Plaintiffs are entitled to partial summary judgment on infringement. 3
1. Ownership and Copyrightability 3
2. Infringement 5
B. Plaintiffs are entitled to summary judgment on Defendants' fair use defenses. 9
1. Factor 1 9
2. Factor 2 17
3. Factor 3 17
4. Factor 4 19
5. Defendants ignore additional uses that must be separately analyzed. 26
C. Plaintiffs are entitled to summary judgment on three elements of their DMCA claims... 27
1. OpenAI's removal was intentional. 27
2. OpenAI removed Plaintiffs' CMI. 28
3. Plaintiffs conveyed CMI in connection with their Asserted Works. 29
D. Plaintiffs are entitled to separate awards of statutory damages for infringement of each article. 30
E. Plaintiffs are entitled to summary judgment on Defendants' affirmative defenses. 31

TABLE OF AUTHORITIES

Page(s)

Cases

Advanta-Star Automotive Research Corp. of America v. Reynolds Ford, Inc.,
2020 WL 5823537 (W.D. Okla. Sept. 30, 2020) 30

American Geophysical Union v. Texaco Inc.,
60 F.3d 913 (2d Cir. 1994).....12, 13

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

Arica Institute, Inc. v. Palmer,
970 F.2d 1067 (2d Cir. 1992).....12

Authors Guild, Inc. v. HathiTrust,
755 F.3d 87 (2d Cir. 2014).....9, 18

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

Bartz v. Anthropic PBC,
787 F.Supp.3d 1007 (N.D. Cal. 2025) 11, 17, 27

Bryant v. Media Right Productions, Inc.,
603 F.3d 135 (2d Cir. 2010).....30, 31

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

Cartoon Network LP, LLLP v. CSC Holdings, Inc.,
536 F.3d 121 (2d Cir. 2008).....9

Celotex Corp. v. Catrett,
477 U.S. 317 (1986).....31

Compania Del Bajo Caroni (Caromin), C.A. v. Bolivarian Republic of Venezuela,
341 F. App'x 722 (2d Cir. 2009) 24

Doe 1 v. GitHub, Inc.,
---F.4th--- 2026 WL 2728464 (9th Cir. Sept. 16, 2026).....28, 29

EMI Christian Music Group, Inc. v. MP3tunes, LLC,
844 F.3d 79 (2d Cir. 2016).....9, 31

Epidemic Sound, AB v. Meta Platforms, Inc.,
2026 WL 2021347 (N.D. Cal. July 13, 2026).....4

Estate of Smith v. Cash Money Records, Inc.,
253 F.Supp.3d 737 (S.D.N.Y. 2017).....13

Feist Publications, Inc. v. Rural Telephone Servs Co.,
499 U.S. 340 (1991).....16

Google LLC v. Oracle America, Inc.,
593 U.S. 1 (2021).....13, 14, 15

Harper & Row Publishers, Inc. v. Nation Enterprises,
471 U.S. 539 (1985).....13, 16, 17, 21

International News Service v. Associated Press,
248 U.S. 215 (1918).....22

Kane v. Comedy Partners,
2003 WL 22383387 (S.D.N.Y. Oct. 16, 2003).....12

Mango v. BuzzFeed, Inc.,
970 F.3d 167 (2d Cir. 2020).....28

MidlevelU, Inc. v. ACI Information Grp.,
989 F.3d 1205 (11th Cir. 2021).....17

Morris v. Business Concepts, Inc.,
283 F.3d 502 (2d Cir. 2002).....3

New York Times Co. v. Tasini,
533 U.S. 483 (2001).....3

Nihon Keizai Shimbun, Inc. v. Comline Business Data, Inc.,
166 F.3d 65 (2d Cir. 1999).....16, 17, 19

Personal Keepsakes, Inc. v. Personalizationmall.com, Inc.,
975 F.Supp.2d 920 (N.D. Ill. 2013).....29

Post University v. Course Hero, Inc.,
2023 WL 5507845 (D. Conn. Aug. 25, 2023).....29

Rogers v. Koons,
960 F.2d 301 (2d Cir. 1992).....8

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

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

Thomson Reuters Enterprise Centr GmbH v. ROSS Intel. Inc.,
193 F.4th 271 (3d Cir. 2026).....*passim*

Wozniak v. Warner Bros. Ent. Inc.,
726 F.Supp.3d 213 (S.D.N.Y. 2024).....4

Statutes

17 U.S.C. § 106.....1

17 U.S.C. §107.....21

17 U.S.C. § 410(c) 3

17 U.S.C. § 504.....30, 31

17 U.S.C. §1202(b)(1) 29

17 U.S.C. §1202(c)(3).....30

Copyright Act.....1, 12, 13, 31

Digital Millennnium Copyright Act.....28, 29

Other Authorities

D. Atkinson, *Our Artificial Future* (2026).....26

TABLE OF CITATION CONVENTIONS

Citation	Document	ECF No.
NP-Mot	News Plaintiffs' Memorandum of Law in Support of Motion for Summary Judgment	1728
NP-SOF	News Plaintiffs' Corrected Rule 56.1 Statement of Material Undisputed Facts	1985
Bauerlein Decl.	Declaration of Monika Bauerlein in Support of News Plaintiffs' Motion for Summary Judgment	1788
Maisel Decl.	Declaration of Jennifer B. Maisel in Support of News Plaintiffs' Motion for Summary Judgment	1815
Goldstein Decl.	Declaration of Tom Goldstein in Support of News Plaintiffs' Motion for Summary Judgment	1819
Taylor Decl.	Declaration of Robert Taylor in Support of News Plaintiffs' Motion for Summary Judgment	1826
Macala Decl.	Declaration of Miroslav Macala in Support of News Plaintiffs' Motion for Summary Judgment	1836
Koonce Decl.	Declaration of Lacy H. Koonce, III in Support of Ziff Davis's Motion for Summary Judgment	1859
OAI-Mot	OpenAI's Memorandum of Law in Support of Motion for Summary Judgment on News Plaintiffs' Claims	1749
OAI-SOF	OpenAI's Rule 56.1 Statement	1735
NP-Opp	News Plaintiffs' Omnibus Memorandum of Law in Opposition to Defendants' Motions for Summary Judgment	2119
NP-CSOF	News Plaintiffs' Counterstatement of Material Facts	2126
Pl-Resp-OAI-SOF	Class and News Plaintiffs' Response to OpenAI's Rule 56.1 Statement	2202
Maisel-Opp-Decl	Declaration of Jennifer B. Maisel in Support of News Plaintiffs' Omnibus Memorandum of Law in Opposition to Defendants' Motions for Summary Judgment	2150

Citation	Document	ECF No.
MSFT-Opp	Defendant Microsoft Corporation's Opposition To News Plaintiffs' Motion For Partial Summary Judgment	2225
MSFT-Class-Opp	Defendant Microsoft Corporation's Opposition To Books Plaintiffs' Motion For Partial Summary Judgment	2173
Simpson Decl.	Declaration of Lisa T. Simpson in Support of Defendant Microsoft's Oppositions to the Books and News Plaintiffs' Motions for Partial Summary Judgment	2201
OAI-Opp	OpenAI's Opposition to News Plaintiffs' Motion for Partial Summary Judgment	2216
Blanco-Opp-Decl	Allison S. Blanco Declaration in Support of OpenAI's Opposition to News Plaintiffs' Motions for Summary Judgment	2183
Krishnan Decl.	Declaration of Sanjay Krishnan, Ph.D. in Support of OpenAI's Opposition to News Plaintiffs' Motion for Partial Summary Judgment	2140
Def-Resp-NP-SOF	Defendants' Response to News Plaintiffs' Rule 56.1 Statement	2253
Def-Daubert-Mot	Memorandum of Law in Support of Defendants' Rule 702 Motion to Exclude Expert Testimony	2052
Pl-Daubert-Opp	Plaintiffs' Opposition to Defendants' Rule 702 Motion to Exclude Expert Testimony	2352
NP-Ex	News Plaintiffs' Exhibits	1777, 1785, 1786, 1790, 1792, 1799, 1803, 1807, 1809, 1812, 1823, 1829, 1830, 1833, 1834, 1837, 1839, 1841, 1842, 1843, 1844, 1846, 1847, 1849, 1851, 1852, 1854, 1856, 1857, 1861, 1863, 1869, 1871, 1874,

Citation	Document	ECF No.
		1876, 1877, 1878, 1880, 1882, 1885, 1887, 1888, 1891, 1893, 1896, 1897, 1899, 1902, 1903, 1905, 1909, 2136, 2148, 2162, 2163, 2165, 2166, 2168, 2172, 2176, 2177, 2178, 2182, 2186, 2193, 2198, 2209, 2214, 2219, 2221, 2223, and those Filed Concurrently with this Brief
NP-Reply-R56.1	News Plaintiffs' Responses to Defendant OpenAI's "Additional Facts"	Filed Concurrently with this Brief

I. Introduction

The unlicensed, commercial copying of millions of original news stories to substitute for them is at the core of what copyright law was designed to prevent. The Copyright Act grants copyright holders the exclusive rights to reproduce, distribute, and create derivatives of their works, 17 U.S.C. § 106, rights that exist to foster human creativity because copying content is cheap and easy but producing it takes sustained investment. Rejecting Defendants' attempt to appropriate these rights for AI companies at the expense of human creators, thereby reducing the incentives for humans to create expressive works, would not "rewrite copyright law" as Microsoft argues, MSFT-Opp-3, but instead reinforce its enduring importance to the continued production of new journalism. Thus, Defendants cannot meet their burden to show fair use.

First, Defendants have not shown that using news to scrape, train, ground, output, or barter with each other lacks substitutive purposes and effects. Indeed, Defendants' position that they can copy entire protected works and provide substitutes, so long as they don't deliver exact copies to their users, is belied by their own cases. *Authors Guild v. Google, Inc. ("Google Books")*, 804 F.3d 202, 218 (2d Cir. 2015) (placing crucial weight on the fact that Google designed its product not to substitute for the books that it enabled users to find). Nor should this Court credit Defendants' efforts to undermine the ample record evidence of existing and future harm, including decreased traffic, consumers' increasing reliance on chatbots for news, and Defendants' own witnesses' predictions that allowing copying here would create a "doom loop" that would result in far less original content being produced. Defendants argue this is not the right kind of harm and that none of this matters for training, because the Court should assess substitutive impact when it trained its LLMs rather than after it incorporated them into public-facing products. But if Defendants' argument were correct, the Supreme Court in *Andy Warhol Foundation of the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508 (2023), would have analyzed substitution when Warhol made his

silkscreen of Prince's portrait—not when the resulting image was licensed for use on a magazine cover.

Second, Defendants fail to address numerous undisputed facts in News Plaintiffs' motion, including OpenAI's successful "hack" of news sources' paywalls, NP-Mot-56, targeting news content for training, NP-Mot-11–13, documents describing memorization and regurgitation as expected products of training, NP-Mot-12, specifically training models to summarize news, NP-Mot-31, and promoting them for that use, NP-Mot-16. These facts show that Defendants' efforts to deliver news were not limited to grounding, as OpenAI claims. Instead, Defendants designed their LLMs to substitute for the journalism on which they were trained, including "evergreen" content that has value long after its initial publication. And Defendants make no serious attempt to justify their unlawful acquisition of publisher content and "horse trading" it with each other, aside from simply trying to lump all three distinct uses together as "training."

Third, in copying News Plaintiffs' journalism to deliver responses to Copilot users, Microsoft doesn't deny that Copilot delivers no meaningful traffic to its grounding sources, NP-Mot-68. Instead, Microsoft depicts News Plaintiffs as sore losers for objecting to the use of their works to compete with them for online readers, MSFT-Opp-3. It's no wonder that among 56 amicus briefs, not one attempts to defend grounding on or output of copyrighted content as fair use.

Beyond fair use, Defendants' nitpicking of a few now-withdrawn works fails to raise a material issue of fact regarding *prima facie* infringement of 3,671,948 works that they copied. Nor are there any factual disputes about OpenAI's removal of copyright management information from News Plaintiffs' works, or, for purposes of statutory damages, about the fact that Defendants copied, and News Plaintiffs license, these works as individual articles.

This case has existential ramifications for a news industry that is vital to the future of our democracy. At bottom, permitting Defendants' conduct as fair use disincentivizes human

expression, and "society would be poorer if the sparks of human creativity become fewer or dimmer" as a result. U.S. Copyright Office, Copyright and Artificial Intelligence, Part 2: Copyrightability, at 37 (January 2025).

II. Argument

A. Plaintiffs are entitled to partial summary judgment on infringement.

1. Ownership and Copyrightability

Plaintiffs' copyright registrations cover all 3,996,365 articles on which summary judgment of infringement is sought and establish *prima facie* ownership and copyrightability. Defendants make no attempt to rebut that showing for 3,671,948 articles[1] aside from arguing that the presumption of ownership does not apply because each article was not separately registered.[2] OAI-Opp-19. But their own cases recognize that copyright in a collective work extends to the constituent parts with the same owner as the whole—as is the case here. *See Morris v. Business Concepts, Inc.*, 283 F.3d 502, 506 (2d Cir. 2002); *New York Times Co. v. Tasini*, 533 U.S. 483, 494 (2001). Plaintiffs' Rule 1006 summaries identify *additional* ownership bases for *each* asserted article. Maisel Decl., Exs. A, B, E; Bauerlein Decl., Ex. A; Macala Decl., Ex. A; Koonce Decl.,

[1] Defendants challenge 317,101 articles: 39 on copyrightability (8-NYT; 7-DNP; 16-Ziff; 8-CIR) and 317,062 on ownership (274,462-NYT; 38,117-DNP; 4,483-Ziff). Def-Resp-NP-SOF-295-432 (de-duplicated to the extent possible, but the vast majority of challenged articles were not specifically identified). This excludes 720,011 DNP articles Dr. Borck challenged based on "institutional bylines" because these bylines prove—not disprove—ownership. *See* Def-Resp-NP-SOF-383; NP-SOF-372-383. More generally, non-specific challenges regarding ownership and copyrightability—offered by Dr. Borck, who has no expertise in copyright—do not undermine Plaintiffs' showing of validity. Additionally, CIR and Intercept inadvertently omitted 7,315 and one article(s), respectively, from their Rule 1006 summaries respectively and withdraw those articles from their motions.

[2] That ZD filed some registrations over five years after first publication goes only to the registrations' weight (17 U.S.C. § 410(c)) but does not negate ownership, and ZD followed the same reliable process for all registrations (Taylor Decl. ¶¶ 19-32). Separately, the testimony OpenAI claims invalidates ZD's database registrations is inapposite; it concerns USCO deposit copies, not the underlying databases. Blanco-Opp-Decl-Ex-252 at 70:19-71:16.

Exs. AA-TT. Because Defendants fail to rebut ownership or copyrightability, summary judgment is appropriate for at least the 3,671,948 unchallenged works. *See Wozniak v. Warner Bros. Ent. Inc.*, 726 F.Supp.3d 213, 227 (S.D.N.Y. 2024) (presumption remains intact for all unchallenged works).

For the remaining 317,101 articles, Defendants contest copyrightability for 39 works and ownership for 317,062 works. *See supra* n.1. Plaintiffs agree to withdraw, for purposes of summary judgment, the 39 copyrightability-challenged articles[3] and 1,949 ownership-challenged articles.[4] Having removed these articles from the dispute, Defendants' remaining arguments fail to raise genuine issues of material fact regarding ownership:

- Seven of Defendants' challenges concern ***other articles*** not asserted in this case. Def-Resp-NP-SOF-327, 373, 384. *Epidemic Sound, AB v. Meta Platforms, Inc.*, 2026 WL 2021347, at *18 (N.D. Cal. July 13, 2026) ("Meta must actually dispute the evidence which Epidemic has put forward, rather than cast doubt on Epidemic's prima facie case based on copyrights not at issue.").

- Defendants identify no relevant differences between online and print Times articles. OAI-Opp-18-19, n.11.

- Plaintiffs' Rule 56.1 statement identifies the witness testimony referenced in the Rule 1006 summaries. *See, e.g.*, NP-SOF-322-384.

- The cited agreements show the chain of title from "Freedom Communications" to named plaintiff ORB Publishing, LLC. NP-SOF-335-338 (concerns 35,416 articles).

- OpenAI's critique of ZD ownership records ignores unrefuted testimony. Taylor Decl. ¶ 40.

- CIR's expert did not include a particular 75-word article in her 200-words-and-over list, as OpenAI claims. NP-Reply-R56.1-1832.

[3] *See* NP-Reply-R56.1-1865-1870 and Krishnan Decl. (identifying 39 articles).

[4] DNP drops 1,875 articles. Def-Resp-NP-SOF-327 (5-articles), 380 (1-article, Blanco-Opp-Decl-Ex.90), 382 (1,868-articles by McDonald, Melvin); Maisel-Opp-Decl. ¶ 5 (1-article). ZD drops 74 articles. Def-Resp-NP-SOF-389 (74-articles by Wopkin, Kavountzis, Hay. *See* Koonce Decl., Ex. BB).

- OpenAI wrongly identifies a 75-word article on CIR's 200-words-and-over list. NP-Reply-R56.1-1832

- The other challenges were immaterial, such as the existence of a third-party contributor (which does not disprove ownership/co-ownership) (e.g., Def-Resp-NP-SOF-380, 383), the lack of a person's name on a byline (other evidence ties bylines to ownership) (e.g., Def-Resp-NP-SOF-320-327, 383), or the absence of testimony that HR records exist in the ordinary course (disclosed witnesses will testify to these facts at trial, and stipulations can resolve this dispute) (e.g., Def-Resp-NP-SOF-320-327).

2. Infringement

Training:

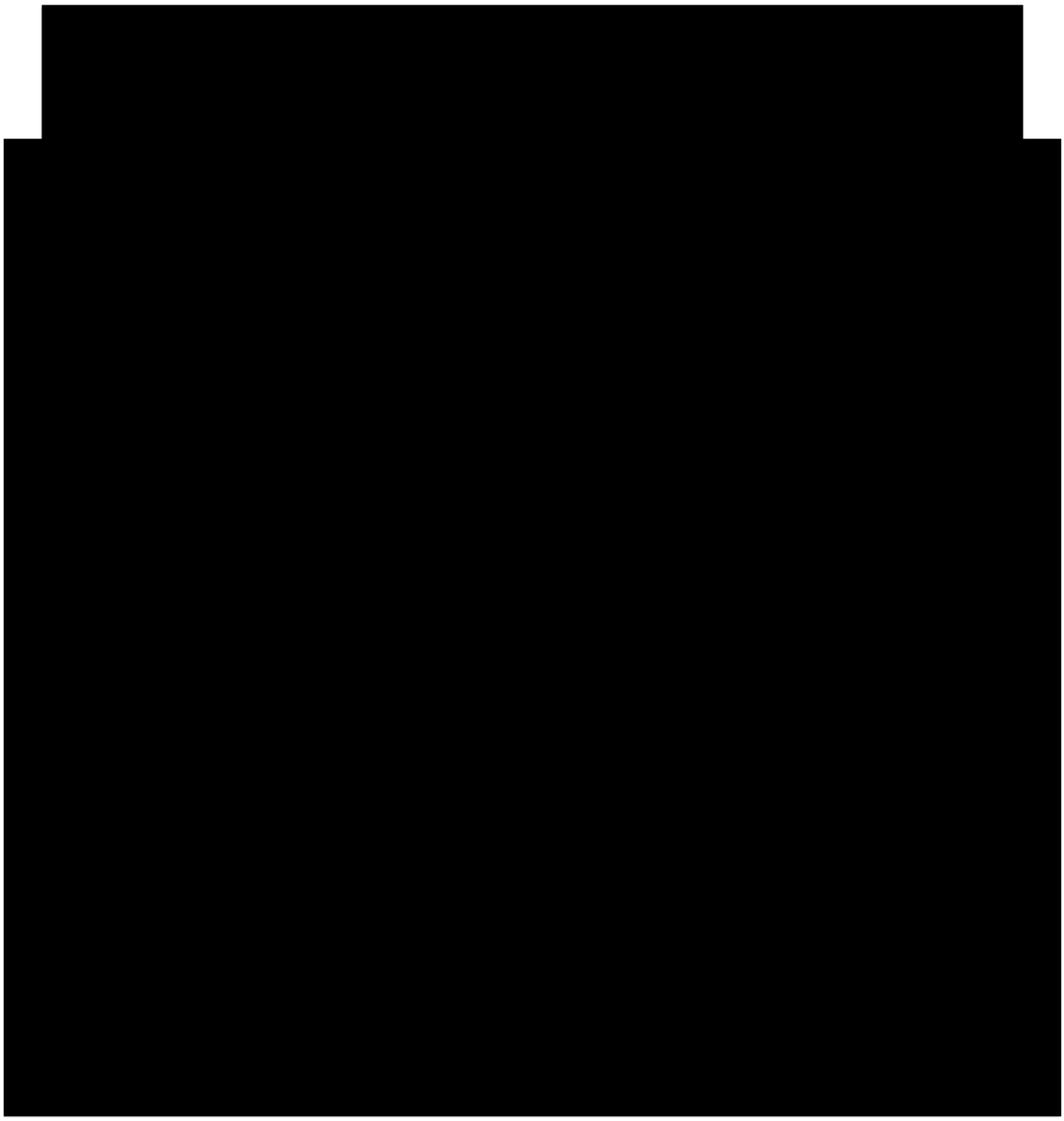
OpenAI criticizes Plaintiffs' experts for using different methods to identify works in the training datasets but fails to identify any false positives. OAI-Opp-21. And because OpenAI does not dispute that it copied the identified works in their entirety, there is no role for the "ordinary observer" test to determine if a copy contains copyrightable elements. OAI-Opp-20-21.

OpenAI asks the Court to conclude that Plaintiffs' comprehensive showing of verbatim copying is flawed based on six isolated examples of third-party text found in Plaintiffs' articles. OAI-Opp-22. However, Plaintiffs took extensive measures to remove potentially unprotectable expression; the fact that OpenAI's own expert could find only six such examples (which have now been removed, *see supra* at page 4) among nearly 4 million works shows those measures' effectiveness. NP-SOF-1227.

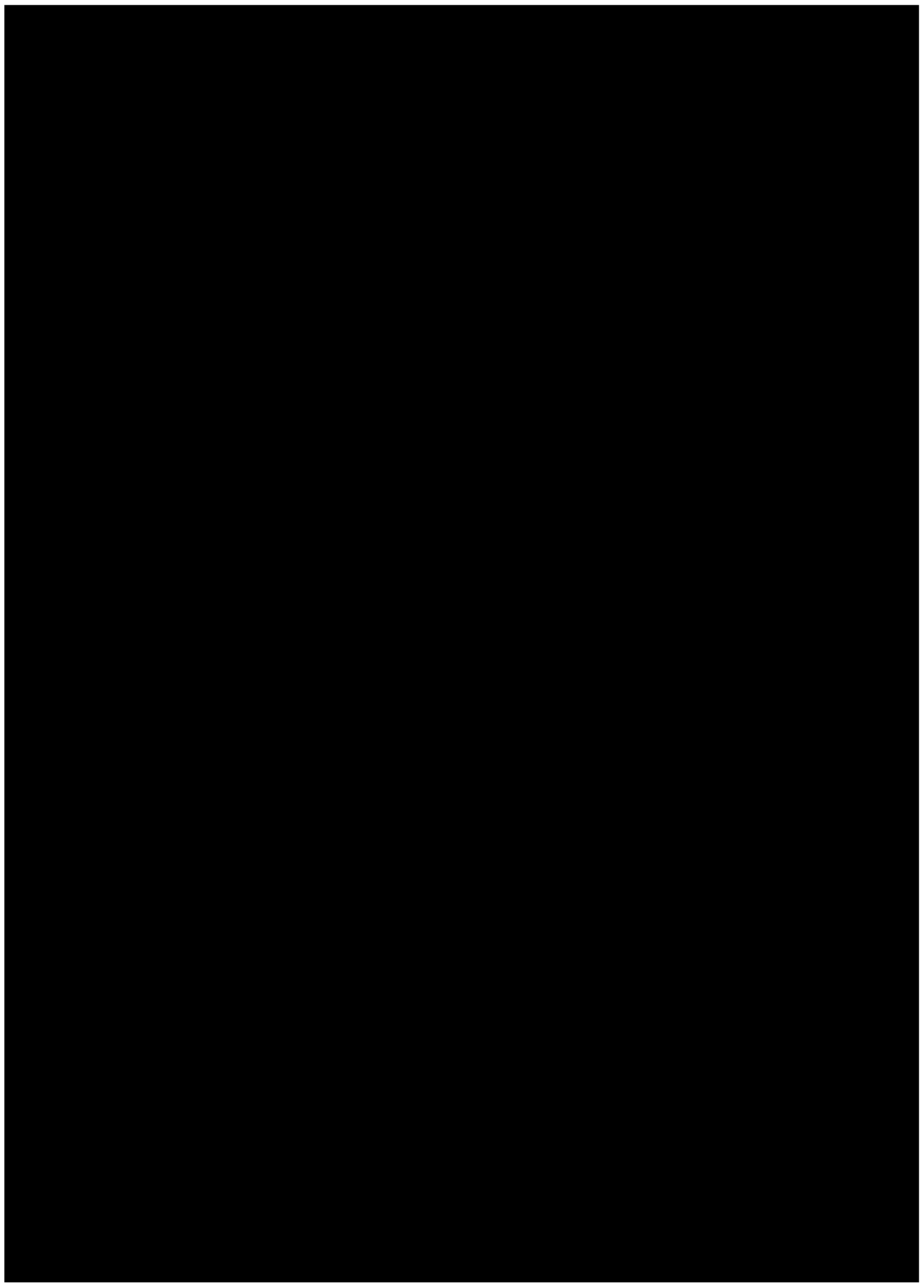
Microsoft RAG/Outputs:

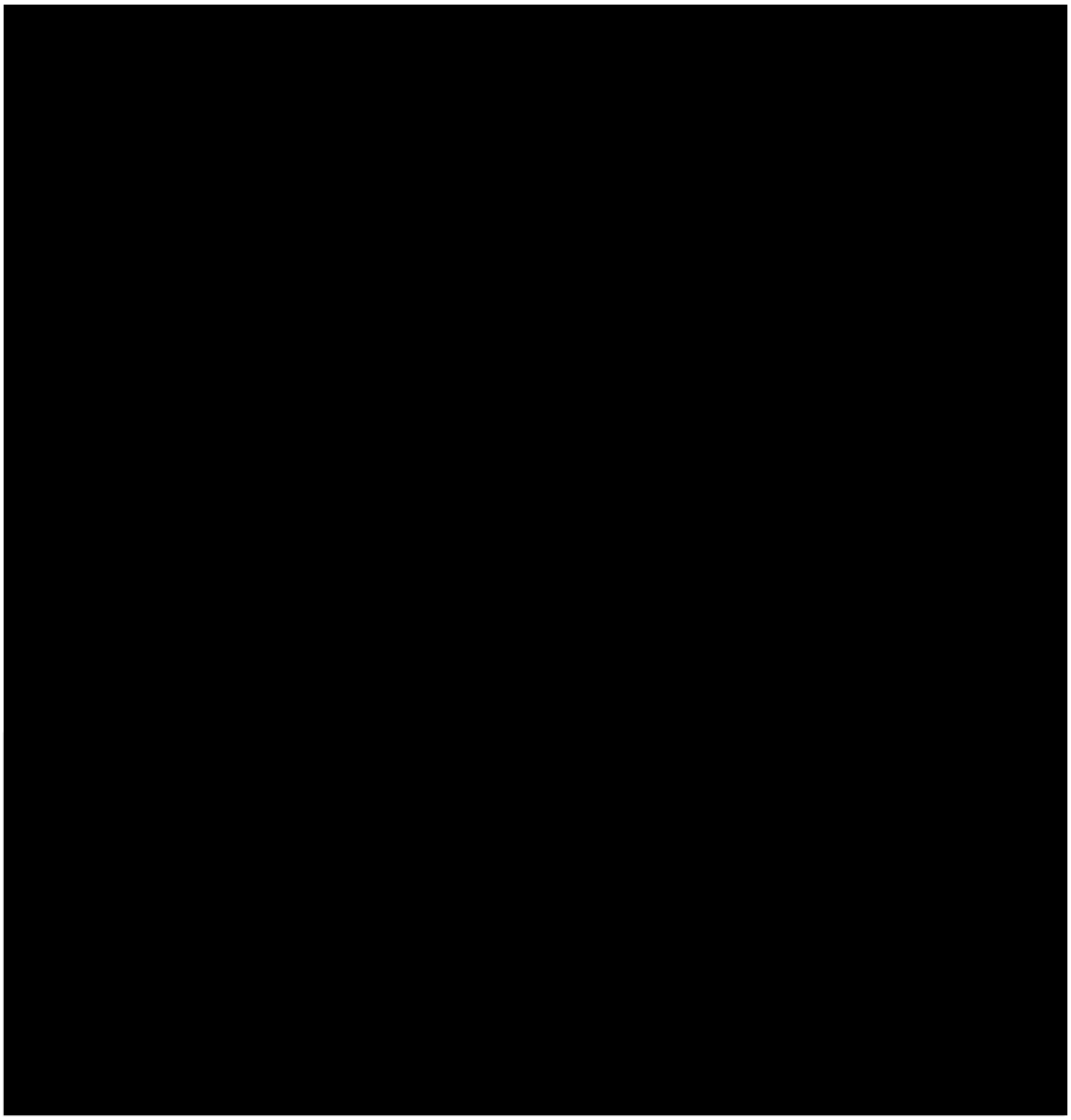
Plaintiffs submitted *prima facie* evidence that Microsoft infringed certain Asserted Works via its RAG pipeline and through Copilot chat outputs. *See* Goldstein Decl., Exs. O and P; NP-Ex-1170. Microsoft argues Plaintiffs did not provide "coverage scores" and "side-by-side comparisons" like Plaintiffs did for training, MSFT-Opp-29, but Plaintiffs' evidence is equivalent to side-by-side comparisons. The only specific evidence Microsoft disputes is a single example among the 68 on which *Daily Plaintiffs* sought summary judgment. MSFT-Opp-30-31. Even that

challenge fails because it ignores the 144 words copied in the output and the additional 513 words copied in the "internal search results" (RAG response) of a 1,487 word article asserted by Chicago Tribune—as the excerpt below shows. *See* Goldstein Decl., Ex. P, Example 23.









Considering the extent of copied text in RAG and output, "no reasonable juror could find that copying did not occur." *Rogers v. Koons*, 960 F.2d 301, 307 (2d Cir. 1992).

Out of the 161 Copilot conversations on which Plaintiffs move for summary judgment as to infringement, Microsoft disputes infringement for three on volitional actor grounds (because a user requested the copyrighted content or input its content into the prompt). MSFT-Opp-27-29.[5]

[5] Microsoft does not address the myriad examples where Copilot reproduced Plaintiffs' content without a request to do so. *See EMI Christian Music Grp., Inc. v. MP3tunes, LLC*, 844 F.3d 79, 96 (2d Cir. 2016); Goldstein Decl., Ex. O, pp.18-21 (accessing and reproducing content from nytimes.com when a user asked about "fallout from the new global economy").

In two of those conversations, Microsoft generated its own query, searched the Internet for related subject matter, evaluated the content from numerous websites, copied the Asserted Work from a different, third-party website, and fed the copied work to an LLM to generate an infringing output. *See* Goldstein Decl., Ex. O, pp. 327-330, 434-37. In the third, Microsoft did the same except that it copied the Asserted Work directly from CIR's website. Simpson Decl. Ex. 348.A at -104. This is a far-cry from a system that "automatically obeys commands and engages in no volitional conduct." *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 131 (2d Cir. 2008).

B. Plaintiffs are entitled to summary judgment on Defendants' fair use defenses.

1. Factor 1

Training:

Defendants cite no evidence connecting training on Plaintiffs' works to purposes other than substituting for news. Substantial evidence shows instead that training on news has just this purpose and effect. NP-Mot-53-54,58; NP-Opp-9. A "transformative" use for purposes of the first fair use factor must not only "serve[] a new and different function from the original work," but also must "not [be] a substitute for it." *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 96 (2d Cir. 2014). In other words, by definition, a use cannot be transformative for purposes of copyright law if it serves as a substitute for the original.

Defendants argue their LLMs serve a "vast multitude" of other purposes, OAI-Opp-7, MSFT-Opp-9, but that is irrelevant under *Warhol*: a use whose objective purpose includes substitution will not tilt the first factor in favor of fair use merely for also being "innovative, in some sense" or "creative progress." 598 U.S. at 543 n.18; *see also Google Books*, 804 F.3d at 218.[6]

[6] Both *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992) and *Sony Computer Ent., Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000) involved copying of non-expressive works to achieve compatibility with functional elements of computer software that bears no relationships to copying news to train LLMs for the reasons explained in Plaintiffs' summary judgment opposition. *See* NP-Opp-9-10 & n.3.

It is not true, as Defendants claim, that their admissions regarding the substitutive effect of their models are just about RAG. OAI-Opp-6. OpenAI policy director Jack Clark's 2020 statement that "our work on AI and Creativity is going to increasingly lead to us creating systems that substitute for the labor of the people that define the 'culture' of society" [redacted], NP-SOF-1677; NP-Ex-1061, and predates OpenAI's introduction of "Browse" in ChatGPT in 2023 by years, NP-SOF-569. So does Dario Amodei's identification of "News Generation" in 2020 as one of GPT-3's "Skills." NP-SOF-1429.

Co-founder Greg Brockman called OpenAI's GPT 3.5 model "very good at any news task" years before Browse was introduced because "every time i do any generative stuff on NYT it seems to predict the next sentence pretty well." NP-SOF-1425. OpenAI contends that the model in question was not trained on the specific articles that Brockman asked it to complete. Def-Resp-NP-SOF-1425. But it does not dispute that it trained that model on hundreds of thousands of Plaintiffs' articles, Def-Resp-NP-SOF-564,799; OAI-Mot-3; Maisel Decl. Ex. A; Koonce Decl. Ex. B, or refute that this training contributed to the model being "excellent at news," Def-Resp-NP-SOF-1425. *See* NP-Opp-4.

OpenAI's Nick Turley did not distinguish between training and RAG when he wrote "our products are largely substitutive, *period*." NP-SOF-1473 (emphasis added). He explained, "our products—even a search product—", *i.e.* not just Browse, "will get more and more substitutive as they get better." NP-SOF-1474. Defendants' claims that regurgitation rates are low, OAI-Opp-8-9, even if credited, do not diminish the significance of their admissions about the substitutive "purpose and character" of the output in connection with the first factor. "[T]he first factor considers" instead "whether and to what extent an original work and secondary use have substitutable purposes," and outputs without regurgitated content still substitute for the original work. *Warhol*, 598 U.S. at 536 n.12. Whether training LLMs on news causes "actual or potential

market substitution," through regurgitation or otherwise, is the focus of the fourth factor, not the first. *Id.*

OpenAI also claims to have required "an enormous text corpus" to train general purpose AI models. OAI-Opp-7. But Defendants do not dispute the truth of their own experts' admissions that such a corpus need not include news. Def-Resp-NP-SOF-857,863,864,867. Dr. Lafferty's opinion that "exposure to news-like text may improve a model's ability to generate news-like … outputs," Def-Resp-NP-SOF-1428, confirms rather than refutes his and Defendants' other experts' testimony that the specific purpose of training on news is to generate text that can substitute for news. NP-Mot-8. And such substitution is not limited to "current events" that post-date model training, OAI-Opp-8, because Plaintiffs' journalism includes "evergreen" content such as investigative reports, product reviews, explainers, health and wellness features, and other topical material, and their online archives serve reader interest in historical articles. *See, e.g.*, NP-CSOF-6-12; NP-SOF-1802-05.

Kadrey and *Bartz* are not persuasive authority that training AI on news is justified under the first factor for reasons already explained, including that the analyses in these cases rely on flawed analogies to human learning and pre-*Warhol* out-of-circuit precedent that conflates purported innovation with transformation. *See* NP-Opp-11-13. OpenAI's attempted distinction of *Meltwater* and *Hachette* on the ground that "[n]either case addressed development of a new technology that can be used to create new, non-infringing works," OAI-Opp-8, mischaracterizes the technology at issue here. LLMs do not generate "works" of human authorship, the only kind of creation the Copyright Act seeks to promote. *See* NP-Opp-14-16. And the distinction addresses a straw man: Plaintiffs do not cite either case on whether intermediate copying for a substitutive purpose tilts the first factor against fair use. Plaintiffs cite *ROSS* and *Texaco*, which both found that

it does. *See* NP-Mot-54, NP-Opp-9; *Thomson Reuters Enter. Ctr. GmbH v. ROSS Intel. Inc.*, 193 F.4th 271 (3d Cir. 2026); *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913 (2d Cir. 1994).

Nor does it matter whether *ROSS* involved "generative" AI. OAI-Opp-9. *Ross* found that fair use did not apply even though, in that case and unlike here, there were no substitutive derivative texts being created. *ROSS* observed that in cases justifying intermediate copying under the first factor, "copying was necessary to access the unprotected functional aspects of computer code." 60 F.3d at 286. ROSS, by contrast, "[did] not need to copy Thomson Reuters's headnotes to access the underlying unprotected information" in judicial opinions to train its AI. *Id.* The "generative" features of Defendants' AI models do not change the undisputed fact that they did not need to copy Plaintiffs' works without permission to train those models.[7]

Defendants cite no case in which the first fair use factor favored a *substitutive* competing secondary use. *See* OAI-Opp-9. The "competing" self-help book in *Arica Institute, Inc. v. Palmer* was a scholarly work of research that commented on and criticized the work from which it borrowed. 970 F.2d 1067, 1077-78 (2d Cir. 1992). *Kane* mocked the original work. *Kane v. Comedy Partners*, 2003 WL 22383387, at *4 (S.D.N.Y. Oct. 16, 2003). Such uses that "target[] an original work may have compelling reason to conjure up the original by borrowing from it." *Warhol*, 598 U.S. at 531. Training lacks such an "independent justification" because the "original work and copying use share the same or highly similar purpose[]" and "wide dissemination of [the] secondary work would otherwise run the risk of substitution for the original or licensed derivatives of it." *Id.*[8]

[7] Further, as Plaintiffs have elsewhere shown, LLMs do not extract uncopyrightable underlying facts from the materials on which they are trained. *See* NP-Opp-9-10.

[8] *Estate of Smith v. Cash Money Records, Inc.* did not find that the song at issue competed with the one that it sampled and grounded its analysis of transformative purpose under the first factor in an

Defendants also do not deny the profit motive underlying even their loftiest claimed objectives: "[redacted]" NP-Ex-1061. Nor do they dispute that a commercial purpose will "weigh against" fair use under the first factor absent a "persuasive transformative purpose." *Google Books*, 804 F.3d at 219. OpenAI quibbles with the characterization of this weight as a "presumption." OAI-Opp-10. But *Campbell* merely rejected a "hard" presumption because the fair uses enumerated in the Copyright Act—including news reporting—"are generally conducted for profit in this country." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 584-85 (1994). This does not override *Harper's* assertion that "commercial use of copyrighted material is presumptively [] unfair" when "the user stands to profit from exploitation of the copyrighted material without paying the customary price." *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985). The presumption of unfairness against a commercial use that, like copying to train or operate an LLM, "consumes" a work as a "factor of production" remains alive and well after *Campbell*. *See* NP-Mot-55; NP-Opp-11; 4 Nimmer on Copyright § 13F.05; *Texaco Inc.*, 60 F.3d at 922.[9]

RAG:

Microsoft claims that RAG is "transformative" because it provides "an up-to-date source of information" to produce "synthesized responses drawn from material across the web" to "engage with the user to better locate, understand, and use information." MSFT-Opp-14. In other words, Copilot's RAG process copies and then conveniently repackages the substance of Plaintiffs' works precisely at the point of consumer demand, keeping those readers for itself. This is the

aesthetic assessment of the copier's motives that *Warhol* supersedes. 253 F.Supp.3d 737, 750 (S.D.N.Y. 2017).

[9] OpenAI incorrectly claims that courts now refuse to consider bad faith in the factor 1 analysis. OAI-Opp-10. The Supreme Court did not overrule *Harper* in *Oracle* and the Second Circuit has not overruled *NXIVM*. Both precedents bind this Court to consider Defendants' piratical approach to acquiring training content. *See* NP-Mot-52.

opposite of transformation. Defendants have not satisfied their "burden to justify [their] taking of [Plaintiffs'] work with some reason other than, 'I can make it better.'" *Warhol*, 598 U.S. at 547 n.21.

Calling Copilot or Browse a "'highly creative and innovative tool,'" MSFT-Opp-15 (quoting *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 30 (2021)), does not address *Warhol* and *Oracle*'s requirement of justification. *See also* NP-Opp-16-17. The use at issue in *Oracle* was justified not because it was innovative or creative, but because "Google used Sun's code in a 'distinct and different' context, and 'only insofar as needed' or 'necessary' to achieve Google's new purpose" of creating a compatible software development platform. *Warhol*, 598 U.S. at 543 n.18 (quoting *Oracle*, 593 U.S. at 30). RAG is nothing like this use, and Defendants do not even attempt to measure RAG against *Warhol's* justification standard.

Microsoft argues that RAG is not a disfavored "consumptive use" of protected works for their intrinsic purpose because "[w]eb grounding uses tokenized, intermediate copies of web content to enable an LLM to generate a new, direct response to a user query based on its parameters and the grounding content in the LLM's context window." MSFT-Opp-15. But this just deploys technical jargon to admit the charge.[10]

RAG so obviously substitutes for its sources that Microsoft devotes considerable space to pretending its "Director of Applied Science," Ph.D. computer scientist Dr. Brent Hecht, could not

[10] Microsoft misleadingly quotes the Nimmer treatise as supporting fair use when it leads to "the creation of [something] new," MSFT-Opp-15 (quoting 4 Nimmer on Copyright § 13F.05) (alteration in original)), but Nimmer's actual quote says it is the creation "of new works" that matters. 4 Nimmer on Copyright § 13F.05. Microsoft presumably altered the quote because its chatbots do not produce "works" in a copyright sense, which requires human expression. See NP-Opp-14-16.

have meant it when he plainly said so. MSFT-Opp-4-5.[11] But "CEO Satya Nadella, who of course *does* speak for Microsoft on AI issues," MSFT-Opp-2, admitted that chatbot conversations substitute for publisher content by "giving you the information right there on the website on the AI platform versus needing to go to the underlying source." NP-SOF-1432.

Microsoft's assertion that Plaintiffs do not "claim substitution for *their protected expression*," MSFT-Opp-16, is false. Plaintiffs have marshalled evidence showing that grounded chatbot conversations siphon demand that would otherwise flow to original expressive news sources, as Mr. Nadella admits. NP-SOF-1432. And just as importantly for the first factor, Plaintiffs claim, and the evidence shows, that the *purpose* of grounding on news is to capture and exploit that demand. Def-Resp-NP-SOF-570,995,1508,1522,1669. Moreover, RAG entails wholesale copying of Plaintiffs' works; Defendants' products cannot, and do not, copy only "facts." NP-CSOF-44. Microsoft cites no evidence to suggest otherwise.

It is also wrong that only substitution in the form of verbatim copying is "cognizable" under the first factor. *ROSS*, 193 F.4th at 284 (factor one weighed against fair use even where defendant "did not use the [copied content] in precisely the same fashion" as plaintiff). Defendants' case regarding their supposed right to copy to "provid[e] direct factual answers in a conversational format," MSFT-Opp-16, concerned humans relaying uncopyrightable information that they

[11] And while Microsoft tries to denigrate Dr. Brent Hecht as not "an executive" and "a divergent" thinker, MSFT-Opp-1, he is in fact a Partner and Director of Applied Sciences at Microsoft. In that role, he [redacted] NP-Ex-671 at 18:10-16; NP-Ex-1373. [redacted] NP-Ex-671 at 22:1-18. Dr. Hecht's work contributed to cross-company efforts, including [redacted] *Id.* at 28. [redacted] *Id.* at 120-24.

learned from an authorized original copy without making another, unauthorized copy of protected expression.[12] Grounding LLMs does not work that way. *See* NP-Opp-13&n.7.

Microsoft's claim that it "is 'entitled … to answer' queries with 'unprotected fact[s]' and 'information'" under *Google Books*, MSFT-Opp-16 (omission in original), chop-quotes a passage from that case's discussion of the fourth fair use factor, not the first. *See* 804 F.3d at 223. The larger context explains that the "cumbersome, disjointed, and incomplete nature of the aggregation of snippets made available through [the] snippet view" at issue made it unlikely "that snippet view could provide a significant substitute for the purchase of the author's book." *Id.* at 224-25. The fluent integration of grounding sources into chatbot conversations, as Mr. Nadella attests, has the opposite purpose. The substitutive impact is particularly acute for news publishers, whose businesses depends on maintaining direct relationships with their readers. NP-SOF-1568.

Microsoft's short quotation of *Harper & Row* to justify "providing direct factual answers in a conversational format" by "copyright's interest in the 'free flow of information,'" MSFT-Opp-16-17 (quoting 471 U.S. at 558), similarly ignores what follows: "'The fair use doctrine is not a license for corporate theft, empowering a court to ignore a copyright whenever it determines the underlying work contains material of possible public importance.'" 471 U.S. at 558. The Court concluded:

> In our haste to disseminate news, it should not be forgotten that the Framers intended copyright itself to be the engine of free expression. By establishing a marketable right to the use of one's expression, copyright supplies the economic incentive to create and disseminate ideas.

[12] *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 348 (1991). And *Nihon* supports Plaintiffs. *See Nihon Keizai Shimbun, Inc. v. Comline Bus. Data, Inc.*, 166 F.3d 65, 73 (2d Cir. 1999) (not fair use because "Comline's abstracts compete with and supersede the Nikkei articles").

Id. The idea that AI should be free to copy news to create a substitute for news that destroys the incentives to produce news is a short-sighted and self-destructive perversion of the doctrine of fair use.

2. Factor 2

Defendants' sole argument on Factor 2, that Plaintiffs' works are "predominantly factual" with only minimal expressive features, ignores the actual texts of the Asserted Works at issue, which include opinion pieces and editorials, in-depth investigative journalism, and Pulitzer-prize-winning articles. NP-CSOF-6-16. Plaintiffs' deeply reported and carefully crafted journalism is not just reporting sports scores or conveying when an event will occur, making it far unlike the particular financial articles at issue in *Nihon*, 166 F.3d 65. *Contra* MSFT-Opp-11. Plaintiffs' storytelling and analysis is the core of what copyright law protects, and it does not fall out of that core simply because it includes facts as well.

3. Factor 3

Training: Defendants concede they trained their LLMs on entire copies of Asserted Works, claiming (1) this was "reasonably necessary" given the volume of content LLM training requires, and (2) that Factor 3 can be ignored where the purpose is "transformative." MSFT-Opp-11; OAI-Opp-11-12. Both points fail.

First, Defendants argue that Plaintiffs confuse ***reasonably*** necessary with ***strictly*** necessary, MSFT-Opp-11; OAI-Opp-11-12—but the record shows Plaintiffs' content was not necessary in ***any*** sense, NP-Mot-58; NP-SOF-863-867. Indeed, Defendants' purported willingness to allow publishers to opt-out of training on their content underscores this very point. OAI-Opp-12 n.9. While Defendants claim *Bartz* and/or *Google Books* foreclose the argument that Defendants need not train on news content, neither case concerned news content, and *Google Books* is further distinguishable since, unlike here, it was "literally necessary" to copy the works to achieve the

alleged transformative purpose. *Google Books*, 804 F.3d at 222. Notably, *ROSS* reached the opposite outcome where, as here, the defendant "took more than necessary." 193 F.4th at 287.

Second, the notion that "[t]he scale and purpose of pretraining LLMs [] amply justified OpenAI's use of millions of complete works"—or, as Microsoft puts it, ***any*** copying is reasonable if the purpose is transformative—lets the fair use exception swallow the rule. OAI-Opp-12; MSFT-Opp-11. Fair use is not a one-factor test. Moreover, the only purpose for which copying news content is potentially necessary is the purpose of delivering news. That is not a transformative use. *HathiTrust*, 755 F.3d at 96 ("[A] transformative work is one that serves a new and different function from the original work and is not a substitute for it.").

RAG: Microsoft argues that RAG involves only "intermediate" use, and that its purported "opt out" mechanism is a get-out-of-jail-free card. MSFT-Opp-17. Neither argument succeeds.

As Plaintiffs' summary judgment opposition explains, "intermediate" copying is neither transformative nor fair where it is performed to create a substitute for the original—regardless of whether the internally copied data is shown to a user. NP-Opp-8–9 (citing *Texaco Inc.*, 60 F.3d at 918 and *ROSS*, 2026 WL 2932557, at *6 & n.7); *see* NP-Opp-9–10 & n.3.

Contrary to Microsoft's characterization, publishers have no reliable means of excluding their content from being used for RAG with Copilot. Microsoft concedes that its "opt out" mechanism, which consists of repurposed metatags, has not proven effective. MSFT-Opp-18. Moreover, other companies use those same metatags for other purposes. NP-Opp-51; NP-SOF-511-516; NP-CSOF-82. Microsoft cannot demand that publishers alter their business practices and relationships with other companies to avoid Microsoft's infringement, particularly where they have

already made clear to Microsoft that they do not want their content used for this purpose. NP-Opp-51.

Outputs: Microsoft claims not that copying in outputs is fair, but that it is so "vanishingly rare" as to be irrelevant. MSFT-Opp-18. The record belies this claim. NP-Opp-36-40. The notion that the Court should "filter[] out" "unprotected factual elements" from a copied output is nonsensical, MSFT-Opp-19—one cannot disaggregate a statement's expression from its meaning. And Microsoft's suggestion that a copied paragraph of a news article is "de minimis as a matter of law," MSFT-Opp-19, misstates the law, which looks to the qualitative significance of the copying, not to some threshold word or paragraph count. *Nihon*, 166 F.3d at 71 (it "is not possible to determine infringement through a simple word count...").

4. Factor 4

Training:

The "fourth factor focuses on actual or potential market substitution." *Warhol*, 598 U.S. at 536 n.12. Because Defendants combined a trained LLM with RAG in Copilot and ChatGPT, any independent analysis of training market substitution would be limited to early days before OpenAI introduced Browse. It is during that time that OpenAI's executives recognized the substitutive nature of their product. *See* Section II.B.1 (Training), *supra*. Moreover, Defendants do not dispute evidence showing that training on news contributes significantly to the ability of models to produce outputs that are potential substitutes for news both with and without grounding on more news. *E.g.*, Def-Resp-NP-SOF-850,873.

Through post-training (which, along with mid-training, is part of the training process), Defendants baked news generation and RAG capabilities into their LLMs. NP-Reply-R56.1-1859. OpenAI's post-training instructed the LLMs to perform tasks such as "[redacted]"/"[redacted]" "[redacted]," "[redacted]," "[redacted]

[redacted]" and "[redacted]." *Id.* Microsoft's Deucalion post-training instructed the same. NP-SOF-921–24.

In mid-training, OpenAI used Plaintiffs' content to "[redacted]" thereby imbuing the LLMs with the latest news. NP-Reply-R56.1-1859. The training data was therefore not "necessarily months [] old" as OpenAI contends (OAI-Opp-6); it included Plaintiffs' articles published up to and post-dating GPT-4 Turbo's initial November 6, 2023 release date. NP-Reply-R56.1-1859. OpenAI is therefore incorrect that its LLMs' ability to "generate text relating to up-to-date factual information on current events arises from Browse." OAI-Opp-13. That ability arises from the training process as well.

Defendants also deride training content as "old," OAI-Opp-14, ignoring undisputed evidence that "evergreen" content—content that is at least 30 days old—on Plaintiffs' websites drives substantial traffic and revenue. Def-Resp-NP-SOF-1804 ([redacted]% of nytimes.com traffic is to evergreen content); ECF2373-1 at 13 (News/Media Alliance) ("evergreen content also plays a critical role as a driver of online attention and traffic"). Product reviews on Wirecutter, for example, or articles on technology, health, gaming, and other topics—key pillars, for instance, of ZD's published content—continue attracting readers seeking explanations and answers long after publication. NP-SOF-55-60; Pl-Resp-OAI-SOF-212. A product trained on that content to answer the same questions competes with those articles for the same readers, and it captures the value of the writing without paying for it. NP-CSOF-24-26,199-201.

Training and grounding are also mutually reinforcing. In the Copilot conversations at issue in this Motion, 86 works were copied both to train and ground the LLM. *Compare* Maisel Decl., Exs.A,B *with id.,* Ex.E. Training on news also influences outputs grounded on news to sound more news-like. NP-CSOF-32-33; ECF2449-1 (Leppert) at 13 (a model trained on newspapers "is more

likely to do well on tasks involving newspapers"). OpenAI employees admit [redacted] NP-SOF-144,976; NP-Ex-132 at 48:17-49:14; NP-Reply-R56.1-1859. Conversely, in what OpenAI called an "[redacted]," publishers who blocked OpenAI from training on their content would "[redacted]." NP-Reply-R56.1-1859.

There is no requirement that Defendants must copy Plaintiffs' works *again* in outputs to recognize market harm bearing such a clear and undisputed causal connection to copying Plaintiffs works. NP-Opp-23-24; *ROSS*, 193 F.4th 271, 278–79. The plain language of the statute addresses "the effect of the use upon the potential market for or value of the copyrighted work." 17 U.S.C. §107. There is no dispute that substitution for original works is an "effect of the use" regardless of whether substitutive text that an LLM generates regurgitates verbatim content of text on which a model is trained.

Even if Defendants were correct, however, when ChatGPT launched, OpenAI employees were aware that models can memorize and regurgitate data and launched Project Giraffe to block verbatim output of content—but only from entities that sued OpenAI. Def-Resp-NP-SOF-955,1151. While the parties dispute the precise rate of regurgitation—including because of OpenAI's destruction of, and tampering with, the output log data, ECF1618—Defendants admit that regurgitation happens. MSFT-Opp.2, OAI-Opp-5. And even purportedly rare occurrences, "when multiplied many times, become in the aggregate a major inroad on copyright that must be prevented." *Harper*, 471 U.S. 539, 569 (1985). Moreover, Defendants only address verbatim copying, and ignore the undisputed evidence that their products output summaries, derivatives, and other content (including hallucinations and pink slime) that infringe and supplant demand for

Plaintiffs' works. NP-Opp-27-29; NP-Mot-16-30,68; NP-SOF-1075-1092,1171,1683-86,1706-09,1740-41.

RAG:

OpenAI does not dispute that RAG harms the market for Plaintiffs' works regardless of training. OAI-Opp-13. Nor does Microsoft carry its burden to show that the up to 93% decline in click-through-rates does not result from its GAI products giving answers. NP-SOF-1530-37; *accord* ECF2325-1 (Digital Content Next) at 5-6 (reporting "26 percent reduction in unique visitors to DCN's members' sites, 18 percent reduction in total views of their sites, and a 16 percent reduction in total visits to their sites," "median decline of 22 percent" of YoY web traffic in 2026, and 5% YoY decline of "total advertising revenue and total overall revenue").

Not one *amicus* argued that RAG was defensible as fair use. This is because it is an unfair business practice to directly compete with the producer of original content using the producer's own (stolen) content, and the Supreme Court proscribed it nearly a century ago in *International News Service v. Associated Press*, 248 U.S. 215, 239-40 (1918); *accord* ECF2452-1 (Jim Albrecht) at 8-9 (ChatGPT is **"not just a competitive news product, it's the perfect news product, the one they've always wanted: one that lets you skip reading all the articles and have your assistant narrate only the parts you're curious about**."); ECF2332-1 at 18 (SAG-AFTRA) ("[RAG] represents direct substitution, because it copies current reporting at the moment a user requests it."); ECF2325-1 at 10 (Digital Content Next).

Other Market Effects:

Defendants do not claim that their products drive subscriptions, traffic, or any form of revenue to Plaintiffs in exchange for copying their content. At the first hearing in this case, Microsoft's lead counsel boasted of this dynamic, explaining that what is "so great" about Microsoft's GenAI search products "is 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." NP-SOF-1511-13. That is, Defendants impede potential subscribers from visiting Plaintiffs' sites. Defendants' documents admit that their products are largely "substitutive, period," "will get more and more substitutive as they get better," and that publishers face an "existential threat" from them. NP-SOF-1473-74.

Defendants do not dispute that traditionally search has been the largest third-party driver of traffic to Plaintiffs' content. Def-Resp-NP-SOF-82,1504,1586,1608. Thus, the reduced click-through-rate will continue to "choke[] off" the "funnel" of new traffic (and revenue) coming to Plaintiffs. NP-SOF-478. Defendants incorrectly claim this is not evidence of harm, or at least the type of harm that this Court should consider. *Meltwater* 931 F.Supp.2d at 554 (considering reduced "clicked-through" rates compared to Google search and concluding "Meltwater News acts as a substitute for news sites").

Defendants' experts also admit that Google's GenAI search product (AI Overviews)—which like Defendants' products combine an LLM with RAG—is causing great harm to publishers. NP-SOF-1781-88. Defendants contend this is irrelevant, but *Hachette* says otherwise. 115 F.4th at 193 (finding harm "self-evident" if "use were to become widespread").

Notwithstanding Defendants' "guardrails," Plaintiffs are experiencing harm. DNP's revenues and referrals are down. NP-SOF-1604-1617. Defendants admitted that Plaintiffs have implemented numerous strategic initiatives that have masked the full extent of harm caused by Defendants. NP-SOF-1625. CIR and The Intercept's respective monthly visitors are down. NP-SOF-1627. ZD has experienced declines in traffic and revenue. NP-SOF-1534,1536,1626; NP-CSOF-198. The Times has seen traffic declines in its core news product. NP-SOF-1802. Plaintiffs' witnesses have testified about the difficulties of competing with entities that are giving away their

own content for free. NP-SOF-1604,1596,1626. Chicago Tribune sent a cease-and-desist letter, and the recipient blamed ChatGPT—which was used to create the content. NP-SOF-1703.

Survey evidence reflects that 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. NP-SOF-1594. Times audience surveys show that, "74% of US news consumers say they use [GenAI] to get news and 93% of US sports media consumers say they use it for sports at the high end." NP-SOF-1578. Among Times readers who use or pay for ChatGPT, approximately 50% said their use of ChatGPT meant they "no longer need news from the New York Times at all" or "only need news from the New York Times some of the time." NP-SOF-1581.

If these surveys left any doubt about reader behavior, [redacted] (NP-Ex-1371)—[redacted], NP-Ex-1372[13]; *see also* NP-Ex-744. Its later reversal of that termination (after Microsoft witnesses were questioned about it in depositions) was plainly litigation-driven. A.G. Sulzberger explained:

> [P]rice-sensitive people, who are much less likely to say where they're going next, would be inclined to take up the free alternatives that have been illegally created by OpenAI and Microsoft.

NP-Ex-3 at 254:15-24. *See also* NP-SOF-1561; NP-SOF-1549; NP-Ex-8 at 189:7-191:8; NP-SOF-1570; NP-SOF-1548; NP-Ex-1255 at 7:19-8:14; Pl-Resp-OAI-SOF-280, 281.

There is also overwhelming and unrebutted evidence of future likely harm—including forecasts of 2026 pageviews for DNPs reflecting declines as high as [redacted]% as compared to 2025 pageviews (for New York Daily News). NP-SOF-1618-1623. The Times estimated its [redacted]

[13] These exhibits are responsive to Defendants' factor four arguments, an issue for which Defendants bear the burden of proof. *See Compania Del Bajo Caroni (Caromin), C.A. v. Bolivarian Republic of Venezuela*, 341 F. App'x 722, 724 (2d Cir. 2009).

[redacted] NP-Ex-1234 at 5; NP-SOF-1549-1550. These forecasts were ignored by Defendants' experts' "trends" analyses. CIR and The Intercept rely on audience and revenue funnels for their success and survival, and cannot build valuable direct relationships when fewer potential readers discover their content as a result of ChatGPT and Copilot's substitution. NP-CSOF-180-193.

As amicus and Nobel laureate Joseph Stiglitz explains: "when AI intermediaries absorb the attention and revenue that once flowed to producers of reliable information, the incentive to produce that reliable information erodes, and the stock of accurate information available to society may shrink over time, even if the technology of processing and transmitting information improves." ECF2438-1 at 6.

Potential Market Effects:

The most direct evidence of an existing and functional licensing market is the [redacted] licenses pursuant to which Defendants pay entities (including news organizations) for, *inter alia*, the use of their content to train and/or ground LLMs. NP-SOF-147-207. The Third Circuit has already recognized the existence of a training market. *ROSS*, 193 F.4th at 288–89. So too have scores of *amici*. *E.g.*, ECF2386-1 at 3 (AI Coalition for Data Integrity); ECF2460-1 at 11 (AI Startups); ECF2359-1 at 5-6 (Copyright Alliance); ECF2325-1 at 16 (DCN); ECF2417-1 at 4 (Human Consent Foundation); ECF2434-1 at 12 (Moritz); ECF2377-1 at 3 (Open Markets Institute); ECF2393-1 at 13-14 (Recording Industry Association of America); ECF2332-1 at 6 (SAG-AFTRA); ECF2456-1 at 4 (William Morris Endeavor Entertainment); ECF2305-1 at 14-15 (Association of American Publishers); ECF2373-1 at 15 (News/Media Alliance).

With respect to Defendants' arguments that "there are enormous—and very probably insurmountable—barriers" to a licensing market, MSFT-Class-Opp-34, Judge Chhabria responded, "If using copyrighted works to train the models is as necessary as the companies say,

they will figure out a way to compensate copyright holders for it." *Kadrey*, 788 F.Supp.3d at 1036; D. Atkinson, *Our Artificial Future* (2026) ("What are we supposed to make of companies that simultaneously claim their technology could cure cancer and solve global warming, yet are utterly incapable of creating a functioning content licensing regime?").

5. Defendants ignore additional uses that must be separately analyzed.

"Horse Trading"

OpenAI ignores Plaintiffs' argument that it was not fair use to copy news articles for "horse trading" datasets with Microsoft. NP-Mot-31-33,75 ("Defendants . . . us[ed] Plaintiffs' content as a form of currency through self-described 'horse-trading' deals."). Microsoft admits this "horse trading" occurred, including the "Project Mango" and "Project Taxi" data "transfer[s]," but claims that fair use applies because "each transfer" was "related to LLM development." MSFT-Opp-10-11; *see also* Def-Resp-NP-SOF-694,695,699,713,716,728. But Microsoft elsewhere concedes that each use of Plaintiffs' content "requires separate analysis." MSFT-Opp-7; *see also Warhol*, 598 U.S. at 533. Even if it were fair use to copy Plaintiffs' articles to train LLMs (and it is not), distributing that content to others—particularly to gain something of value in exchange—violates Plaintiffs' statutory rights and requires a distinct fair use justification. Defendants have none. Copying Plaintiffs' content for horse trading does not "add[] something new" to the copied works, *Warhol*, 598 U.S. at 528, nor is distributing unlicensed datasets a "new and different way[] of using news articles," MSFT-Opp-11. News publishers already license news articles for training and/or RAG. Def-Resp-NP-SOF-148,207,229,244,247-274.

Acquiring Content to Build Repositories

Defendants scraped and downloaded copies of Plaintiffs' works to build repositories and datasets for many uses. NP-Mot-8-11. OpenAI ignores this distinct use of Plaintiffs' content, while Microsoft lumps it together with its "training" arguments. MSFT-Opp-7. Copying to train cannot

justify the separate practice of copying articles to store them "forever" in a general-purpose repository. NP-SOF-822. Defendants admit they acquired Plaintiffs' content by disregarding their sites' terms of service or scraping their content from "behind a paywall" Def-Resp-NP-SOF-463,521; OAI-Opp-3-4. And it is undisputed that OpenAI "stores" this copied content within its "Azure storage account," where OpenAI has "retained" the content so it can "recreate datasets" in the future. Def-Resp-NP-SOF-813,820,822,844,889,892; *see also* OAI-Opp-3. As in *Bartz*, "[p]irating copies to build a research library without paying for it, and to retain copies should they prove useful for one thing or another, was its own use — and not a transformative one." *Bartz v. Anthropic PBC*, 787 F.Supp.3d 1007, 1026 (N.D. Cal. 2025).

Finally, it is undisputed that OpenAI "used" the New York Times Annotated Corpus to "filter[]" Common Crawl and compile a "dataset" for GPT-3—an LLM that OpenAI admits was used to power its "commercial[]" API product. Def-Resp-NP-SOF-582,585,1119. OpenAI thus violated the "*commercial* use prohibition in the User Agreement for the New York Times's Annotated Corpus" (OpenAI's words), and summary judgment is warranted for the 928,872 copied Times articles from that dataset. Def-Resp-NP-SOF-1102-07,1119 (emphasis added); NP-Mot-10-11,39.

C. Plaintiffs are entitled to summary judgment on three elements of their DMCA claims.

1. OpenAI's removal was intentional.

OpenAI leads with the same flawed arguments as its affirmative motion and is wrong for reasons Plaintiffs explained in response: OpenAI's text extractors targeted CMI-specific fields, its

employees singled out CMI as data to remove, and its removal of CMI for a "purpose" proves that it was intentional. *See* OAI-Opp-24; OAI-Mot-23-24; NP-Opp-51-53; NP-SOF-1303-1359.

2. OpenAI removed Plaintiffs' CMI.

First, OpenAI argues that text extraction does not involve "removal," OAI-Opp-24-25, ignoring the myriad statements by its employees and experts equating them, NP-SOF-1303-1304,1346-1359. And, its only record citation is nonexistent: its Statement of Facts contains nothing resembling the phrase "materially different and entirely distinct extracted text." OAI-Opp-25 (purportedly quoting OAI-SOF-341-346).

That the text extractors "transform raw HTML" does not defeat removal. OAI-Opp-25. Any removal "transform[s]" preexisting content: it deletes its CMI. And that is what OpenAI did: it deleted CMI from copies of Plaintiffs' works it downloaded from the internet. While OpenAI's process involved creating new CMI-less copies, OAI-Opp-25, OpenAI cites no authority that the new copy negates removal. Accepting OpenAI's position would render the DMCA effectively unenforceable: a defendant need simply make a new CMI-less infringing copy. A statute that "strengthen[s] copyright protection in the digital age," *Mango v. BuzzFeed, Inc*., 970 F.3d 167, 170-71 (2d Cir. 2020), cannot possibly incentivize its own evasion through more unlawful copying.

OpenAI's only case, *Doe 1 v. GitHub, Inc*., ---F.4th--- 2026 WL 2728464 (9th Cir. Sept. 16, 2026), did not consider removal of CMI "at the training stage," where OpenAI removed Plaintiffs' CMI. *Id.* at *3. Instead, it held omitting CMI from outputs was not removal because, due to LLMs' "probabilistic" nature, outputs are new content. *Id*. at *7. But OpenAI's extractors operate on "existing work[s]" containing Plaintiffs' CMI. *Id.* at *5. Indeed, text extraction, per OpenAI's expert, "retain[s]" the existing work. Def-Resp-NP-SOF-1354. *GitHub* supports

Plaintiffs by holding that "removal" includes reproducing an existing work while omitting its CMI. 2026 WL 2728464, at *5.

Second, OpenAI argues Plaintiffs' experts failed to reliably identify CMI removals. OAI-Opp-25. This rehashes its flawed *Daubert* arguments. *See* Def-Daubert-Mot-12,22,25; Pl-Daubert-Opp-14-15,22. OpenAI also claims it retained some CMI categories in its training data's "URL" field. OAI-Opp-25. Even if it retained URLs, OpenAI still removed CMI from the webpage, which suffices. 17 U.S.C. §1202(b)(1) (prohibiting removal of "any [CMI]"). OpenAI offers no authority that the DMCA requires removing *all* CMI.[14]

3. Plaintiffs conveyed CMI in connection with their Asserted Works.

First, OpenAI argues that information in Plaintiffs' footers, headers, and navigation bars (some of the claimed CMI) is not sufficiently "connect[ed] to the asserted work." OAI-Opp-25. But CMI can be "accessible in conjunction with, or appear with, the work," which this information is and does. *Post Univ. v. Course Hero, Inc*., 2023 WL 5507845, at *4 (D. Conn. Aug. 25, 2023) (quoting *Janik v. SMG Media, Inc*., 2018 WL 345111, at *12 (S.D.N.Y. Jan. 10, 2018)); NP-SOF-1257-1283. Regardless, anyone would "understand the information to be referring to" the work, *id.*, when the webpage is a news website and the work is a news article, even if publication dates sometimes differ. OpenAI's case is distinguishable: the website belonged to a gift item producer, while the works were poems, generating no similarly obvious inference. *Pers. Keepsakes, Inc. v. Personalizationmall.com, Inc*., 975 F.Supp.2d 920, 922 (N.D. Ill. 2013). OpenAI's case has been

[14] OpenAI also argues CIR relied on training data from third-party sources; however, its two examples are not asserted at summary judgment. OAI-Opp-25; Def-Daubert-Mot-21; NP-Ex-1173.

distinguished on this ground. *Advanta-Star Auto. Rsch. Corp. of Am. v. Reynolds Ford, Inc.*, 2020 WL 5823537, at *4 (W.D. Okla. Sept. 30, 2020).

Second, OpenAI argues NYT/DN/ZD's experts inaccurately identify CMI. OAI-Opp-26. That is wrong. *See* Pl-Daubert-Opp-11-13,25. OpenAI disputes, for example, that a column name can be CMI, but ignores the statute's inclusive language. *Id.*; NP-Reply-R56.1-1842.

Third, OpenAI argues that NYT/DN/ZD's experts wrongly include publisher identity as a CMI category. OAI-Opp-26. But for these Plaintiffs, publisher identity is synonymous with copyright owner, which is valid CMI. *See* 17 U.S.C. §1202(c)(3) (CMI includes "[t]he name of, and other identifying information about, the copyright owner of the work"). OpenAI offers no legal support for its argument, and provides no example where publisher identity is not CMI. OpenAI's challenge is purely theoretical.

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

The articles in this case were issued in two ways: individually in digital form and in a print compilation. The record is undisputed that Defendants copied the individual, digital copies, not the print compilations. *See* OAI-Opp-26-27; MSFT-Opp-31-32.

Statutory damages exist to help rights holders receive fair and effective compensation, and to further policy goals of deterrence. *Bryant v. Media Right Prods., Inc.*, 603 F.3d 135, 144 (2d Cir. 2010). Defendants and others like them have not been deterred. If one could, for example, copy an entire issue of The New York Times and pay only a single statutory damages award, that would not be much of a deterrent. It is therefore appropriate, under both the statute and binding Second Circuit precedent, to award statutory damages on a per article basis.

Defendants argue that Plaintiffs fail to address the statute (17 U.S.C. § 504) and *Bryant*. The statute squarely supports per-article statutory damages because the Defendants scraped and ingested—and then copied, recopied, and distributed—articles (not issues). It states that a

copyright owner may recover statutory damages "for all infringements involved in the action, with respect to any one work…" Defendants infringed articles, not issues. *See* OAI-Opp-26-27; MSFT-Opp-31-32. Further, because Defendants did not infringe the issues (or compilations), the language at the end of § 504—regarding all parts of a compilation constituting one work—is inapplicable.

Bryant also supports per-article damages. *Bryant* explains that "[t]he Copyright Act allows only one award of statutory damages for any 'work' infringed." 630 F.3d at 140. *Bryant* also explains that, for compilations, the question of whether statutory damages are available per individual-work vs. compilation depends on how the infringed works were issued. Were they issued "separately, or together as a unit"? *Id*. at 141. Here, their digital versions were issued separately.

Defendants also argue that *EMI Christian* requires a plaintiff to have ***sold (for money)*** the individual articles in order to recover per article damages. Not so. The focus is on whether the copyright holder ***issued (not sold)*** the works separately, or together as a unit. *EMI Christian*, 844 F.3d at 101. In *Twin Peaks Productions, Inc. v. Publications Intern, Ltd.*, for example, the television episodes were issued sequentially, at different times, but not sold, and statutory damages were awarded per episode. 996 F.2d 1366, 1381 (2d Cir. 1993). In any event, Plaintiffs do license their articles individually. NP-SOF-433-462.[15]

E. Plaintiffs are entitled to summary judgment on Defendants' affirmative defenses.

Defendants failed to identify evidence sufficient to establish the essential elements of those defenses, as required by *Celotex Corp. v. Catrett*, 477 U.S. 317, 322–25 (1986). Defendants

[15] OpenAI claims ZD has not produced copyright certificates for 180,011 works and cannot seek statutory damages for them, but the registrations were identified in ZD's pleading; while ZD did not have certificates in its files, it has obtained them from the USCO and produced them. NP-Reply-R56.1-1864. ZD does not seek statutory damages on its remaining Asserted Works used in training the models currently at-issue.

improperly attempt to shift their burden of proof to Plaintiffs, who are not required to disprove defenses on which Defendants bear the ultimate burden of proof. *See id*.

Dated: November 20, 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@susmangodrey.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 (*pro hac vice*)
Jenny L. Colgate (*pro hac vice*)
Alexandra Hughes (*pro hac vice*)
Mark Rawls (*pro hac vice*)
Michael H. Jones (*pro hac vice*)
Kristen J. Logan (*pro hac vice*)
Bryan B. Thompson (6004147)
Mary L. Mullins (*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
mjones@rothwellfigg.com
klogan@rothwellfigg.com
bthompson@rothwellfigg.com
mmullins@rothwellfigg.com

Attorneys for Daily Plaintiffs

/s/ *Jon Loevy*
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/ *Lance 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.

CERTIFICATE OF WORD COMPLIANCE

Pursuant to Local Civil Rule 7.1(c), the foregoing Memorandum of Law complies with the Court's order of August 7, 2026 (ECF 1664), as it contains 8,740 words, according to the word processing program used to prepare it.

/s/ Ian Crosby
Ian Crosby