UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE: OPENAI, INC., 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	25-md-3143 (SHS) (OTW) Hon. Sidney H. Stein Hon. Ona T. Wang ORAL ARGUMENT REQUESTED PUBLIC VERSION - REDACTED

OPENAI'S OPPOSITION TO NEWS PLAINTIFFS'

MOTION FOR PARTIAL SUMMARY JUDGMENT

TABLE OF CONTENTS

I. INTRODUCTION 1

II. BACKGROUND 3

A. OpenAI's Data Collection and Training for its LLMs 3

B. News Plaintiffs' Allegations of Harm 4

C. News Plaintiffs' Allegations Regarding "Regurgitation" in LLM Outputs 5

III. LEGAL STANDARD 5

IV. ARGUMENT 5

A. News Plaintiffs Are Not Entitled to Summary Judgment on Fair Use 5

1. News Plaintiffs Conflate Training with Browse 6

2. Factor One: OpenAI's Alleged Use Was Highly Transformative. 7

a. OpenAI's Use Serves a Different Purpose than the Originals 7

b. Commerciality Does Not Outweigh Transformativeness 10

c. "Bad Faith" Is Disputed and Immaterial to OpenAI's Defense 10

3. Factor Two: The Asserted Works Are Published and Highly Factual 10

4. Factor Three: OpenAI's Alleged Use Was Reasonable in Amount 11

5. Factor Four: News Plaintiffs Have Not Suffered Market Harm 12

a. OpenAI's Alleged Use Did Not Substitute for News Plaintiffs' Protected Expression and Caused No Harm 12

b. News Plaintiffs Have Not Suffered Cognizable Harm to a Licensing Market 15

c. Widespread Adoption of the Challenged Use Would Not Cause Cognizable Harm 17

B. News Plaintiffs Are Not Entitled to Summary Judgment on Their Claim of *Prima Facie* Infringement 18

1. The Record Does Not Permit Summary Judgment on the Validity of News Plaintiffs' Copyrights 18

a. News Plaintiffs Have Failed to Establish Ownership 18

b. News Plaintiffs Failed to Establish Copyrightability 20

2. News Plaintiffs' Purported Evidence of Substantial Similarity Is Insufficient to Grant Summary Judgment for Millions of Works 20

C. News Plaintiffs' DMCA Claim Fails as a Matter of Law 23

1. News Plaintiffs Cannot Establish Requisite Intent 24

2. News Plaintiffs Cannot Establish CMI Removal 24

3. News Plaintiffs Have Not Reliably Identified Their Own CMI 25

D. Statutory Damages 26

1. News Plaintiffs Are Not Entitled to Statutory Damages on a Per-Article Basis 26

2. Ziff Davis Is Entitled to No Statutory Damages 27

E. News Plaintiffs Are Not Entitled to Summary Judgment on OpenAI's Affirmative Defenses 27

V. CONCLUSION 28

TABLE OF AUTHORITIES

Page(s)

CASES

Advanced Analytics, Inc. v. Citigroup Glob. Markets, Inc.,
301 F.R.D. 31 (S.D.N.Y.), *objections overruled*, 301 F.R.D. 47 (S.D.N.Y. 2014) 22

American Geophysical Union v. Texaco, Inc.,
60 F.3d 913 (2d Cir. 1994).....16

Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith,
598 U.S. 508 (2023).....7, 9, 10

Arica Institute v. Palmer,
970 F.2d 1067 (2d Cir. 1992).....2, 9

Associated Press v. Meltwater U.S. Holdings, Inc.,
931 F.Supp.2d 537 (S.D.N.Y. 2013).....8

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

Authors Guild v. HathiTrust,
755 F.3d 87 (2d Cir. 2014).....6, 17

Bartz v. Anthropic PBC,
787 F.Supp.3d 1007 (N.D. Cal. 2025) 1, 7, 12, 15

Beaulier v. Meta Platforms, Inc.,
2026 WL 2521848 (N.D. Cal. Aug. 26, 2026) 24

Bill Graham Archives v. Dorling Kindersley Ltd.,
448 F.3d 605 (2d Cir. 2006).....15

Blanch v. Koons,
467 F.3d 244 (2d Cir. 2006).....11

Bryant v. Media Right Prods., Inc.,
603 F.3d 135 (2d Cir. 2010).....26

Bustamante v. KIND, LLC,
100 F.4th 419 (2d Cir. 2024) 27

Campbell v. Acuff-Rose Music, Inc.,
510 U.S. 569 (1994).....10, 11, 12, 17

Celotex Corp. v. Catrett,
477 U.S. 317 (1986).....5

Doe 1 v. GitHub, Inc.,
— F.4th —, 2026 WL 2728464 (9th Cir. Sept. 16, 2026).....24, 25

Eldred v. Ashcroft,
537 U.S. 186 (2003).....2

EMI Christian Music Group v. MP3tunes, LLC,
844 F.3d 79 (2d Cir. 2016).....26, 27

Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,
499 U.S. 340 (1991).....20

Fischer v. Forrest,
968 F.3d 216 (2d Cir. 2020).....23

Fox News Network, LLC v. TVEyes, Inc.,
124 F.Supp.3d 325 (S.D.N.Y. 2015), *rev'd on other grounds*, 883 F.3d 169
(2d Cir. 2018).....6

Google LLC v. Oracle Am., Inc.,
593 U.S. 1 (2021).....10, 15, 17

Hachette Book Group, Inc. v. Internet Archive,
115 F.4th 163 (2d Cir. 2024).....8, 11, 14, 16

Hoehling v. Universal City Studios, Inc.,
618 F.2d 972 (2d Cir. 1980).....20

Hofheinz v. AMC Prods., Inc.,
147 F.Supp.2d 127 (E.D.N.Y. 2001).....16

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

Kane v. Comedy Partners,
2003 WL 22383387 (S.D.N.Y. Oct. 16, 2003), *aff'd* 98 F.App'x 73 (2d Cir.
2004).....9

McGucken v. Shutterstock, Inc.,
166 F.4th 361 (2d Cir. 2026).....24

Morris v. Bus. Concepts, Inc.,
283 F.3d 502 (2d Cir. 2002).....19

Murphy Med. Assocs., LLC v. Cigna Health & Life Ins. Co.,
2026 WL 1652610 (D. Conn. May 8, 2026).....20

New York Times Co. v. Microsoft Corp.,
777 F.Supp.3d 283 (S.D.N.Y. 2025).....21

New York Times Co. v. Tasini,
533 U.S. 483 (2001).....19

NXIVM Corp. v. Ross Institute,
364 F.3d 471 (2d Cir. 2004).....10, 12, 13

Pers. Keepsakes, Inc. v. Personalizationmall.com, Inc.,
975 F.Supp.2d 920 (N.D. Ill. 2013).....25

Pierson v. Infinity Music & Ent., Inc.,
300 F.Supp.3d 390 (D. Conn. 2018).....25

Sega Enters. Ltd. v. Accolade, Inc.,
977 F.2d 1510 (9th Cir. 1992), *as amended* (Jan. 6, 1993).....8

Shine v. Childs,
382 F.Supp.2d 602 (S.D.N.Y. 2005).....20

Estate of Smith v. Cash Money Records, Inc.,
253 F.Supp.3d 737 (S.D.N.Y. 2017), *aff'd* 799 F.App'x 36 (2d Cir. 2020).....9

Sony Comput. Ent., Inc. v. Connectix Corp.,
203 F.3d 596 (9th Cir. 2000).....2, 8

Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P.,
756 F.3d 73 (2d Cir. 2014).....11

Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P.,
861 F.Supp.2d 336 (S.D.N.Y. 2012).....20

Thomson Reuters Enter. Centre GmbH v. ROSS Intel. Inc.,
— F.4th —, 2026 WL 2932557 (3d Cir. Sept. 29, 2026).....9, 16

Ushodaya Enters., Ltd. v. V.R.S. Int'l, Inc.,
64 F.Supp.2d 352 (S.D.N.Y. 1999).....27

Vermont Teddy Bear Co. v. 1-800 Beargram Co.,
373 F.3d 241 (2d Cir. 2004).....18

Walker v. Time Life Films, Inc.,
784 F.2d 44 (2d Cir. 1986).....21

White v. West Publ'g Corp.,
29 F.Supp.3d 396 (S.D.N.Y. 2014) 17

Zalewski v. Cicero Builder Dev., Inc.,
754 F.3d 95 (2d Cir. 2014).....18

STATUTES

17 U.S.C.
§ 107(1).....2, 7
§ 107(2).....10
§ 107(4).....2, 12
§ 412.....27
§ 504(c) 26
§ 1202(b)(1) 23
§ 1202(c) 25, 26

RULES

Fed. R. Civ. P. 56(a) 5

OTHER AUTHORITIES

Compendium of U.S. Copyright Office Practices, 3d ed., § 727.....20

I. INTRODUCTION[1]

OpenAI's use of news articles to pretrain its LLMs was fair use as a matter of law. OpenAI trained its models on billions of textual works collected from the public Internet to distill broad patterns of language and facts about the world, so they could generate entirely new text without republishing protected expression from the training data. This use is "quintessentially transformative." *Bartz v. Anthropic PBC*, 787 F.Supp.3d 1007, 1021–22 (N.D. Cal. 2025); *Kadrey v. Meta Platforms, Inc.*, 788 F.Supp.3d 1026, 1044 (N.D. Cal. 2025). And it caused News Plaintiffs no harm. They identify no lost sales, subscribers, or ad revenues. They emphasize a handful of instances in which they claim that OpenAI's LLMs "regurgitated" portions of their articles, but their own experts scoured millions of real-world user chats and found that less than 0.00011% contained any amount of "regurgitation."

Faced with this reality, News Plaintiffs argue that pretraining is not fair use because OpenAI's LLMs can generate "news" or other factual information that competes with news that News Plaintiffs have published or will publish in the future (alleged "market dilution"). They speculate they will suffer future market harm because users will turn to ChatGPT for "fast, timely, answers" instead of going to News Plaintiffs' websites. MDL ECF 1728 ("News Mot.") at 68–69. They admit they have "no empirical evidence" to support this speculation, but claim that "common sense suffices." *Id.* at 68.

News Plaintiffs' theory of harm fails as a matter of law. As a preliminary matter, the fact that OpenAI's LLMs can provide "fast, timely answers" based on up-to-date information arises not from pretraining, but from a distinct "web browsing" feature, which News Plaintiffs' motion

[1] In all quotations appearing herein, all internal quotation marks, citations, alterations, and the like have been omitted, and all emphases added, unless otherwise noted.

does not challenge. MDL ECF 1749 ("OpenAI-News Mot.") at 12–15 & n.7; News Mot. at 12.[2] Pretraining does not enable retrieval of information published after a model's knowledge cutoff, which is at least several months before a model is released. Responses to News' SUF ("RNSUF") ¶ 1859. Whether OpenAI's use of works for pretraining is a fair use depends on the "purpose" and "effect" of that "use," not the purpose and effect of a distinct web-browsing use. 17 U.S.C. § 107(1), (4).

Moreover, facts, including facts that News Plaintiffs deem "news," are not copyrightable and "become[] instantly available for public exploitation at the moment of publication." *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003); *see also Authors Guild v. Google, Inc*. ("*Google Books*"), 804 F.3d 202, 224 (2d Cir. 2015) (holding that lost sales due to the provision of "historical fact[s]" is irrelevant to fair use). Moreover, a use can be fair even where the defendant's product satisfies a consumer need that the plaintiff hopes to satisfy. *Arica Institute v. Palmer*, 970 F.2d 1067, 1078 (2d Cir. 1992) (holding a defendant is "perfectly entitled to create a competing work"); *see also Sony Comput. Ent., Inc. v. Connectix Corp.*, 203 F.3d 596, 607–08 (9th Cir. 2000) (holding that "some economic loss" caused by "competition" from the defendant's transformative product "does not compel a finding of no fair use").

News Plaintiffs also ask the Court to take the unprecedented step of finding that they have established ownership, registration, and infringement of millions of articles as a matter of law, and so are entitled to hundreds of billions in statutory damages. They also seek summary judgment on parts of their DMCA claims, demanding hundreds of billions more in statutory

[2] News Plaintiffs only seek summary judgment on their training claims and "reserve" the right to seek summary judgment on their "grounding and output" claims against OpenAI until after resolution of their baseless sanctions motion. News Mot. at 3 n.1. News Plaintiffs never sought court approval for this approach and it ignores the Court's deadlines. MDL ECF 1664.

damages based on mischaracterized, immaterial facts and arguments contrary to controlling law. On both fronts, they have failed to meet their burden.

For the reasons explained in OpenAI's motion for summary judgment and below, the Court should grant summary judgment in favor of OpenAI and deny News Plaintiffs' motion.

II. BACKGROUND[3]

A. OpenAI's Data Collection and Training for its LLMs

Although immaterial to OpenAI's fair use defense, News Plaintiffs make several misrepresentations regarding OpenAI's data collection and training process for its LLMs.

Use of news content. OpenAI did not specifically target News Plaintiffs' articles for use in training. *Contra* News Mot. at 5–33, 58 n.7. OpenAI sought to train a "generalist" LLM based on a diverse collection of text from myriad Internet sources. News was simply one of many types of text OpenAI used. *See* MDL ECF 1718 ("OpenAI-Class Mot.") § II.D; OSUF ¶¶ 23, 138–139.

Robots.txt. OpenAI did not "sidestep[]" robots.txt by using data collected by Common Crawl. *Contra* News Mot. at 10. Common Crawl maintained its own robots.txt compliance policy, and News Plaintiffs did not block its crawler during the relevant time period. RNSUF ¶ 1861. OpenAI took additional steps to avoid training on News Plaintiffs' content via a blocklist, which automatically filtered text collected from their domains from training datasets. *See* RNSUF ¶¶ 1847–1851.

Paywalls. Nor have News Plaintiffs shown that OpenAI intentionally "circumvent[ed] paywalls and violat[ed] terms of use when acquiring [training] data." *Contra* News Mot. at 9. The fact that content may under certain conditions appear behind a paywall in a standard browser

[3] OpenAI incorporates by reference its summary-judgment papers, including its Statement of Undisputed Material Facts, which provide the relevant background regarding OpenAI's fair use defense. *See* OpenAI-News Mot. at 2–6; MDL ECF 1735 ("OSUF").

does not establish that OpenAI bypassed a paywall to obtain it. OpenAI's data acquisition scripts worked through an industry-standard mechanism (which News Plaintiffs themselves use) whereby OpenAI would send requests to servers and collect the content the servers sent back in response. RNSUF ¶¶ 1853–1858. If that process resulted in the collection of content that may fall behind a paywall if viewed via a standard internet browser, the publishers' servers returned that content in response to OpenAI's automated requests. A publisher's decision to display a paywall to browser users does not establish that the same restriction governed automated requests—or that OpenAI circumvented any restriction. *Id.* And human-readable terms of use are distinct from technical access controls for web crawlers; their presence does not show that OpenAI bypassed a paywall. *Id.*

NYT Annotated Corpus. News Plaintiffs' claim that OpenAI used The New York Times Annotated Corpus to fine-tune GPT-3, MDL ECF 1985 ("NSUF") ¶¶ 1117–1118, is wrong; GPT-3 was not fine-tuned at all. RNSUF ¶¶ 1117–1118. The cited evidence instead concerns OpenAI's permitted use of the Corpus for an unreleased internal research model, which this Court has excluded from the scope of this case. MDL ECF 707 at 1.[4]

Bloom Filter. OpenAI implemented an output filter ("Bloom filter") as an additional measure to prevent "regurgitation" of certain copyrighted content. *See* OpenAI-Class Mot. at 12. News Plaintiffs' assertion that use of their articles in the Bloom filter is itself an infringement ignores this undisputed fact. *See* News Mot. at 13, 39.

B. News Plaintiffs' Allegations of Harm

It is undisputed that The Times has more revenue, subscribers, and page views than before ChatGPT's launch, OSUF ¶¶ 275–277, and pre-ChatGPT traffic trends have not changed

[4] The In-Scope Models are GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o mini. MDL ECF 707 at 1.

for any News Plaintiff. Empirical analyses by OpenAI's expert demonstrate that user adoption of ChatGPT had no negative impact on News Plaintiffs' web traffic. OpenAI-News Mot. at 5–6, 10–11; OSUF ¶¶ 241–243, 247–249, 253, 258–264. News Plaintiffs do not offer any empirical evidence of their own and instead claim that "common sense suffices." News Mot. at 68.

C. News Plaintiffs' Allegations Regarding "Regurgitation" in LLM Outputs

News Plaintiffs do not seek summary judgment against OpenAI based on allegedly infringing outputs, News Mot. at 3 n.1, but nevertheless spend 13 pages discussing purported "regurgitations." These isolated examples do not establish widespread "regurgitation," as News Plaintiffs imply. OpenAI's expert identified only ***24*** instances of verbatim reproduction in a sample of 20 million ChatGPT conversation logs, a rate of 0.00012%. OSUF ¶ 219. News Plaintiffs' own experts' analyses showed similarly low "regurgitation" rates: 0.00011% from The Times/DNP (*id.* ¶ 220); 0.000008% from CIR (*id.* ¶ 222); 0.000002% from The Intercept (*id.* ¶ 221); and 0.00002% from Ziff Davis (*id.* ¶ 224). *See also* OpenAI-News Mot. 4–5.

III. LEGAL STANDARD

Summary judgment is warranted where "the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law." Fed. R. Civ. P. 56(a). The moving party bears the initial burden of demonstrating the absence of a genuine issue of material fact for trial. *Celotex Corp. v. Catrett*, 477 U.S. 317, 323 (1986).

IV. ARGUMENT

A. News Plaintiffs Are Not Entitled to Summary Judgment on Fair Use

OpenAI's alleged use of the Asserted Works to pretrain its LLMs was highly transformative and caused no harm to the market for or value of the Asserted Works.

1. News Plaintiffs Conflate Training with Browse

News Plaintiffs move for summary judgment of no fair use based only on OpenAI's alleged training but improperly rely on evidence concerning OpenAI's web browsing service ("Browse"), which is a separate technology and use.[5] *See supra* note 1. For example, News Plaintiffs claim that ChatGPT's ability to access and discuss "news" about current events in response to a user query means that OpenAI's use of their works for pretraining is not fair. *See, e.g.*, News Mot. at 53–54, 65–71. But that capability is not a result of pretraining, which relies on content that is necessarily months (if not years or decades) old. OpenAI-News Mot. at 10–11; OSUF ¶¶ 142, 208–213. Rather, that capability arises exclusively from Browse, which News Plaintiffs do not challenge in their motion and which is a fair use for reasons explained in OpenAI's News Motion. OpenAI-News Mot. at 12–15.[6] Any purported harm from ChatGPT's ability to discuss breaking news does not arise from pretraining and is irrelevant here.[7]

[5] OpenAI-News Mot. at 1, 12 & n.7.

[6] As explained in the Opposition to Class Plaintiffs' Motion, while courts do not apply the fair-use analysis to each alleged copy, they do conduct separate analyses for distinct "uses" or functionalities. *See Authors Guild v. HathiTrust*, 755 F.3d 87, 97–104 (2d Cir. 2014) (analyzing "Full-Text Search" and "Print-Disabled" uses separately); *Google Books*, 804 F.3d at 216–18, 220–23 (same for "Search Function" and "Snippet View"); *Fox News Network, LLC v. TVEyes, Inc.*, 124 F.Supp.3d 325, 333 (S.D.N.Y. 2015), *rev'd on other grounds*, 883 F.3d 169 (2d Cir. 2018) (same for "four complementary features").

[7] News Plaintiffs' arguments regarding CustomGPTs, News Mot. at 5–7, 16, 20, 67, also fail. CustomGPTs allow a third party to build specialized versions of ChatGPT according to their own, not OpenAI's, specifications. RNSUF ¶¶ 1014–1018. The aspects of CustomGPTs that News Plaintiffs complain about do not arise from OpenAI's alleged use of News Plaintiffs' articles to train models, and News Plaintiffs dismissed with prejudice their contributory infringement claims premised on alleged direct infringement by end users. *See* MDL ECF Nos. 1590-1, 1598-1, 1606, 1608-1, 1661–1662.

2. Factor One: OpenAI's Alleged Use Was Highly Transformative.

a. OpenAI's Use Serves a Different Purpose than the Originals

The first fair use factor considers "the purpose and character of the use." 17 U.S.C. § 107(1). "A use that has a further purpose or different character is said to be 'transformative.'" *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 529 (2023).

OpenAI's alleged use of News Plaintiffs' works was transformative because it served a purpose fundamentally different from the purpose of the Asserted Works. News Plaintiffs wrote their articles to engage human readers. OpenAI's alleged use of those articles in pretraining was to distill broad patterns of human language and facts about the world from an enormous text corpus, with the goal of developing LLMs that could generate entirely new text for a vast multitude of purposes. OpenAI-News Mot. at 7–8.[8]

The only two courts to address whether use of copyrighted content to train a generative AI model is fair use both found that the use was transformative as a matter of law. *Kadrey* held that "[t]here is no serious question that Meta's use of the plaintiffs' books had a 'further purpose' and 'different character' than the books—that it was highly transformative." 788 F.Supp.3d at 1044. *Bartz* held that Anthropic's use of copyrighted books to train LLMs was "spectacularly" and "quintessentially" transformative and described the technology as "among the most transformative many of us will see in our lifetimes." 787 F.Supp.3d at 1021–22, 1033.

News Plaintiffs barely mention pretraining and fail to address *Kadrey* or *Bartz* on this point. Instead, they argue that OpenAI's use cannot be transformative because one potential use of OpenAI's LLMs is to generate new, ***non-infringing*** "news" content. News Mot. at 53–54. As

[8] News Plaintiffs also allege that OpenAI used a subset of the Asserted Works at other stages of the training process. News Mot. at 11–13 (discussing "mid-training" and "post-training"). They make no argument that the fair use analysis is any different for those articles than for pretraining.

noted in Section IV.A.1, however, the fact that OpenAI's LLMs can access and discuss current events has nothing to do with the training use that News Plaintiffs' motion challenges. Moreover, this Court has already held that "abridgements" of news articles that summarize their underlying facts are noninfringing. *N.Y. Times*, No. 1:23-cv-11195, ECF 514 at 40–42.

Regardless, the use of copyrighted content to develop a new technology can be transformative, even if the technology results in increased competition for the plaintiff. *See Google Books*, 804 F.3d at 222–24; *Sega Enters. Ltd. v. Accolade, Inc.,* 977 F.2d 1510, 1522–23 (9th Cir. 1992), *as amended* (Jan. 6, 1993); *Connectix Corp.*, 203 F.3d at 606–07; *see also* OpenAI-Class Opp. at 27–29 (explaining why Class Plaintiffs' similar argument fails).

None of News Plaintiffs' cases suggest otherwise. News Mot. at 52–55. Each involved defendants republishing plaintiffs' expression for the same purposes as the originals. In *Associated Press v. Meltwater U.S. Holdings, Inc.*, the "central feature" of the defendant's product was to "capture and republish" copyrighted text from plaintiffs' news articles for Meltwater's subscribers to read. 931 F.Supp.2d 537, 552 (S.D.N.Y. 2013). In *Hachette Book Group, Inc. v. Internet Archive*, the defendant distributed free digital copies of entire books to be read, the "same exact purpose as the originals." 115 F.4th 163, 181–82 (2d Cir. 2024). Neither case addressed development of a new technology that can be used to create new, non-infringing works.

It is undisputed that OpenAI's LLMs were not designed to republish protected expression from the Asserted Works. News Plaintiffs point to a handful of outputs that show alleged "regurgitation," but their own experts' analyses of millions of real-world user conversations reveal just how rare such outputs are: ranging from 0.000002% (The Intercept) to 0.00011% (The Times/DNP). OSUF ¶¶ 220–224. Such rare instances of small amounts of "regurgitation"

do not change OpenAI's transformative use into a "substitutive" use. *Google Books*, 804 F.3d at 223–25.

Thomson Reuters Enter. Centre GmbH v. ROSS Intel. Inc. does not dictate a contrary result. There, the Third Circuit found ROSS's use "minimally transformative, at best" because both parties used the works to "create and optimize a legal-research platform." — F.4th —, 2026 WL 2932557, at *6 (3d Cir. Sept. 29, 2026). Here, the parties' purposes were fundamentally different. *See* OpenAI-News Mot. 20–25. Indeed, the Third Circuit expressly distinguished OpenAI's use of copyrighted content to develop generative AI capable of "generat[ing] original expression." *ROSS,* 2026 WL 2932557, at *6 n.7.

Nor does *ROSS* dictate that increased competition precludes transformativeness. Such a rule would be irreconcilable with Second Circuit precedent finding transformative uses even where the parties' products compete. *See, e.g.*, *Arica*, 970 F.2d at 1077–78 (competing self-help books); *Kane v. Comedy Partners*, 2003 WL 22383387, at *4 (S.D.N.Y. Oct. 16, 2003), *aff'd* 98 F.App'x 73 (2d Cir. 2004) (competing comedy TV shows); *Estate of Smith v. Cash Money Records, Inc.*, 253 F.Supp.3d 737, 749–50 (S.D.N.Y. 2017), *aff'd* 799 F.App'x 36 (2d Cir. 2020) (competing songs).

News Plaintiffs also rely on internal OpenAI and Microsoft documents to characterize OpenAI's use as "substitutive." News Mot. at 52–54. But those documents do not address substitution for protected expression. *See Google Books*, 804 F.3d at 223; *Warhol*, 598 U.S. at 531–32; *see also* OpenAI-Class Opp. at 15–21. They discuss broad LLM capabilities, including performing tasks traditionally done by humans. Those capabilities underscore the transformative nature of OpenAI's use. *Kadrey*, 788 F.Supp.3d at 1044 (LLMs can "generate diverse text and perform a wide range of functions"). In any event, the first-factor requires "an objective inquiry

into what use was made," not OpenAI's "subjective intent" regarding the use (and certainly not the intent of third parties in using OpenAI's products). *See Warhol*, 598 U.S. at 533, 544–45.

b. Commerciality Does Not Outweigh Transformativeness

News Plaintiffs wrongly assert that OpenAI must "overcome [a] presumption" against fair use because its alleged use was commercial. News Mot. at 55. The Supreme Court has held that commerciality carries no "presumptive force" against fair use, *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 584–85 (1994), and should instead "be weighed against the extent to which the use is transformative or otherwise justified." *Warhol*, 598 U.S. at 537 n.13. And the Second Circuit has "repeatedly rejected the contention that commercial motivation should outweigh a convincing transformative purpose and absence of significant substitutive competition with the original." *Google Books*, 804 F.3d at 219.

c. "Bad Faith" Is Disputed and Immaterial to OpenAI's Defense

News Plaintiffs accuse OpenAI of "transgressive and deceptive" data collection, supposedly undertaken "to avoid the cost of acquiring lawful copies." News Mot. at 56. But accusations of "bad faith … contribute[] little to fair use analysis." *NXIVM Corp. v. Ross Institute*, 364 F.3d 471, 478–79 & n.2 (2d Cir. 2004); *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 32–33 (2021) (expressing "skepticism about whether bad faith has any role in a fair use analysis"). Worse, News Plaintiffs grossly distort the record through mischaracterized documents and selectively-quoted deposition testimony as detailed in OpenAI's Rule 56.1 counterstatement. *See* RNSUF. But even if undisputed, their allegations could not outweigh OpenAI's highly transformative purpose. *See Google*, 593 U.S. at 32–33.

3. Factor Two: The Asserted Works Are Published and Highly Factual

The second factor considers "the nature of the copyrighted work[.]" 17 U.S.C. § 107(2). News Plaintiffs do not dispute that the Asserted Works are published and substantially factual,

both of which favor fair use. *See* News Mot. at 57; *Campbell*, 510 U.S. at 586; *Blanch v. Koons*, 467 F.3d 244, 256 (2d Cir. 2006). Instead, News Plaintiffs incorrectly argue that "[f]actor two turns on whether the works ... reflect the author's individualized expression." News Mot. at 57. But the second fair use factor is focused not on whether the works at issue contain some amount of creative expression—which is a requirement for a work to be copyrighted—but rather how creative and thus how close to the "core" of copyright they are. *Campbell*, 510 U.S. at 586; *Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P.*, 756 F.3d 73, 87, 89 (2d Cir. 2014).

If copyrightable at all, many of the Asserted Works lie far from the core of copyright protection. *See* OpenAI-News Mot. at 26–27. In any event, News Plaintiffs' own authorities acknowledge that factor two "rarely play[s] a significant role" in fair-use analysis. *Hachette*, 115 F.4th at 187.

4. Factor Three: OpenAI's Alleged Use Was Reasonable in Amount

Factor Three asks whether the amount copied was "reasonable in relation to the purpose of the copying." *Campbell*, 510 U.S. at 586–87; OpenAI-News Mot. at 9. "The relevant consideration ... is not the amount of copyrighted material used by the copier, but 'the amount of copyrighted material made available to the public.'" *Hachette*, 115 F.4th at 188; *see also Google Books*, 804 F.3d at 221–22. OpenAI's alleged use of complete works was reasonable given its transformative purpose: pretraining its LLMs required hundreds of billions to tens of trillions of tokens from diverse sources, OSUF ¶¶ 23, 37–38, and OpenAI did not make its training data publicly available, *id.* ¶ 34.

News Plaintiffs argue that "Defendants did not need [their] content to ... build[] LLMs capable of generalization." News Mot. at 58. This argument misstates the law. The question is not whether it was strictly or technically necessary to use any one work or portion of a work. Rather, the relevant question is whether use of a complete work was "***reasonable*** in relation to

the purpose of the copying." *Campbell*, 510 U.S. at 586–87. News Plaintiffs present no argument that OpenAI's use was unreasonable and they once again fail to address *Bartz* and *Kadrey*, both of which held that use of entire books for pretraining was reasonable. *Bartz*, 787 F.Supp.3d at 1030; *Kadrey*, 788 F.Supp.3d at 1050. Indeed, *Bartz* rejected the argument News Plaintiffs make here: because the volume of text required to train LLMs is "monumental," "using any one work for actually training LLMs was about as reasonable as the next" and so "'reasonably necessary' does not mean 'strictly necessary.'" 787 F.Supp.3d at 1030; MDL ECF 1748 ¶¶ 74–77. The scale and purpose of pretraining LLMs thus amply justified OpenAI's use of millions of complete works.[9]

5. Factor Four: News Plaintiffs Have Not Suffered Market Harm

Factor Four examines the effect of the use on the "market for or value of the copyrighted work." 17 U.S.C. § 107(4). The "only harm" that matters is harm from substitution for the protected expression of the original work. *Campbell*, 510 U.S. at 593; *Google Books*, 804 F.3d at 223–24. "[T]he relevant market effect with which we are concerned is the market for plaintiffs' 'expression,' and thus it is the effect of defendants' use of that expression on plaintiffs' market that matters, not the effect of defendants' work as a whole." *NXIVM*, 364 F.3d at 482.

News Plaintiffs offer no evidence rebutting OpenAI's showing (OpenAI-News Mot. at 7–12) that pretraining caused no cognizable market harm. They identify no lost subscriptions, no lost sales, and no loss in ad revenue attributable to OpenAI's alleged use.

a. OpenAI's Alleged Use Did Not Substitute for News Plaintiffs' Protected Expression and Caused No Harm

OpenAI's alleged use of the Asserted Works does not substitute for their protected

[9] Moreover, in 2023, OpenAI introduced the use of robots.txt as a ready way for publishers to exclude works that appear on their websites from collection by OpenAI for pretraining LLMs. *See* OSUF ¶ 176; NSUF ¶¶ 484, 488.

expression because OpenAI's LLMs do not display a meaningful amount of that expression to users. Rather, OpenAI's LLMs generate new, non-infringing text.

News Plaintiffs do not seriously dispute this point. Although they claim that OpenAI's LLMs "produce outputs that contain verbatim copies of Plaintiffs' asserted works as well as paraphrased copies and detailed summaries," News Mot. at 65, their experts concede that "regurgitation" is exceedingly rare. *See supra* Section II.C. News Plaintiffs do not claim they have suffered actual market harm from such a tiny amount of "regurgitation." Instead, they claim that ChatGPT substitutes for their publications by providing factual information they characterize as "news." *See* News Mot. at 65–71. This theory fails for at least three independent reasons.

First, as noted above, the ability of OpenAI's LLMs to generate text relating to up-to-date factual information on current events arises from Browse, not pretraining. News Plaintiffs do not challenge Browse in their motion, and thus their theory fails at the outset.

Second, copyright does not protect facts, and competitive "harm" from republishing factual information is not cognizable market harm. *Google Books*, 804 F.3d at 224; *accord NXIVM*, 364 F.3d at 482; *see also* OpenAI-Class Opp. § III.B.4.a (discussing "market dilution").[10]

Third, even if competition from non-infringing "news" were cognizable, News Plaintiffs offer no evidence rebutting OpenAI's showing that ChatGPT caused no loss in subscriptions, web traffic, or revenues. They assert that "empirical evidence" of harm is unnecessary because

[10] News Plaintiffs complain about so-called "pink slime" journalism that they contend is generated using "AI." If News Plaintiffs' complaint is that AI can help small companies produce content that competes with them, that argument fails for all the reasons described above. If News Plaintiffs mean to suggest that "pink slime" websites plagiarize their articles, News Mot. at 69–70, that is a complaint against third parties, not OpenAI.

"common sense suffices." *See* News Mot. at 60, 68. But this argument misapplies *Hachette*. *See id.* at 60–61. The defendant in *Hachette* created a "competing substitute" because it republished entire books to users for free. 115 F.4th at 189–90. In that specific context, the court made the "logical inference" that consumers would be unlikely to pay for copyrighted content they could get for free. *Id.* at 189–90, 193. Here, News Plaintiffs ask the Court to "infer" that using stale news articles to train a language model somehow displaced demand for their publications. That inference makes no sense. Pretraining does not make training data available to readers. And articles used in pretraining would have been at least months old by the time the resulting model became available to users. *See* OSUF ¶¶ 143, 208–13. Even if users were turning to ChatGPT to learn the facts contained in these old articles, that would have zero impact on demand for present-day breaking news. *Hachette* does not stand for the proposition that the court must always draw all inferences in plaintiff's favor on market harm, especially where, as here, OpenAI has presented unrebutted evidence that News Plaintiffs have not suffered cognizable harm. *See Kadrey*, 788 F.Supp.3d at 1057 (declining to draw such an inference; distinguishing *Hachette*).

News Plaintiffs' theory that ChatGPT deprives them of website traffic and potential subscriptions fares no better. News Mot. at 67–68. This theory again depends on "Browse," not pretraining. Moreover, even if users were flocking to ChatGPT for "fast, timely answers," *id.* at 69, that would not show cognizable harm: ChatGPT generates new text rather than republishing the Asserted Works. OpenAI "would be entitled, without infringement of [News Plaintiffs'] copyright, to answer" user questions using factual information derived from the Asserted Works (or any other work). *See Google Books*, 804 F.3d at 224. Copyright law does not guarantee News Plaintiffs a certain volume of website traffic or "clicks." In any event, their analysis measures

click-through rates only on Microsoft properties, not ChatGPT. News Mot. at 67–69.

Nor do The Times's internal surveys show market harm. *See* News Mot. at 70–71. At most, they show that some Times subscribers say they use ChatGPT, not that ChatGPT caused a decline in Times' web traffic or subscriptions, let alone that ChatGPT did so because it displaced demand for The Times's protected expression. *See id.* The same goes for News Plaintiffs' carefully worded assertion that DNP and ZD have "experienced declining search referrals and revenue since ChatGPT launched in 2022." *Id.* at 70. Any such declines predated ChatGPT. OpenAI-News Mot. at 5–6; OSUF ¶¶ 240–288. OpenAI's expert demonstrated that ChatGPT did not cause any change to those pre-existing trends, and News Plaintiffs conducted no empirical analysis showing otherwise. *See id.*

b. News Plaintiffs Have Not Suffered Cognizable Harm to a Licensing Market

News Plaintiffs argue that OpenAI harmed an existing or likely-to-be-developed licensing market by using the Asserted Works without paying. News Mot. at 71–72. But a copyright owner cannot establish market harm simply by charging others for the challenged use. OpenAI-News Mot. at 11 & n.5; OpenAI-Class Mot. at 36–38. The Supreme Court warned against the "danger of circularity posed" by treating the potential market as "the theoretical market for licensing the very use at bar." *Google*, 593 U.S. at 38.

Use of text to pretrain LLMs is transformative, and alleged harm from a "lost" licensing fee in a "transformative market" is not cognizable. *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 614–15 (2d Cir. 2006); *see also Kadrey*, 788 F.Supp.3d at 1052 (hypothesized market "is not one that the plaintiffs are legally entitled to monopolize"); *Bartz*, 787 F.Supp.3d at 1032 (market "is not one the Copyright Act entitles Authors to exploit").

News Plaintiffs ignore these authorities and instead rely on cases involving non-

transformative uses. *See* News Mot. at 61–62, 71–72. In *American Geophysical Union v. Texaco, Inc.*, the defendant photocopied scientific articles for the same purpose as the originals—the only question was whether there was a market for individual articles rather than complete journals in which articles traditionally appeared. 60 F.3d 913, 925 (2d Cir. 1994). In *Hachette*, the defendant redistributed complete copies of books for reading. 115 F.4th at 181. And in *ROSS*, the defendant used Westlaw headnotes to build a legal research platform, which was the same purpose for which Westlaw used them, including to train its own AI search products. 2026 WL 2932557, at **6, 9.

Even if News Plaintiffs' licensing theory were legally cognizable, it would fail because there is no traditional or likely-to-be-developed market for licensing news for pretraining. OpenAI's LLMs are pretrained on hundreds of billions to tens of trillions of tokens from diverse sources. OSUF ¶ 23. Any given work is an infinitesimally small part of the whole, and the transaction costs for licensing individual works (or groups of works) would vastly exceed the marginal benefit of those works. *Id.* ¶¶ 23–29; *see also* OpenAI-Class Mot. at 36–38. The partnership agreements News Plaintiffs cite, News Mot. at 71–72, do not show otherwise. They center on materially different rights, such as the right to display verbatim content from partner websites, and/or were motivated by the formation of a broader collaborative partnership with another party. NSUF ¶¶ 147–150; MDL ECF 1857-1 ¶¶ 26, 179–183, 191–198; *Google Books*, 804 F.3d at 226–27 (rejecting reliance on licensing agreements involving materially different uses of the works). They do not establish the existence of a viable market for the pretraining use at issue here. *See id.*; *see also Hofheinz v. AMC Prods., Inc.*, 147 F.Supp.2d 127, 139 (E.D.N.Y. 2001) (rejecting argument that "defendants paid to exhibit clips of shorter duration than those at issue in the infringement claims," because there is "an equally persuasive inference [] that

defendants sought to avoid litigation"). In any case, "it is irrelevant that [some AI developers] might be willing to purchase licenses in order to engage in this transformative use[.]" *HathiTrust*, 755 F.3d at 100. "Lost licensing revenue counts under Factor Four only when the [accused] use serves as a substitute for the original," *id.*, which is not the case here.

Finally, News Plaintiffs wrongly suggest that licensing costs would be modest. Their own damages expert estimated training license fees of $0.25 to $2.68 per word. MDL ECF 1800, Ex. A at 144, Fig. 35; *id.*, Ex. F at 101, Fig. 20. Applying News Plaintiffs' own estimates to the volume of text needed to train a model comparable to [redacted] would suggest licensing costs of $2 trillion to $26 trillion—before accounting for transaction costs. *See* OSUF ¶ 26 ("[redacted]."); MDL ECF 1863-23 ¶ 180 (assuming that "each word roughly equates to 1.33 tokens"). If recognized, the licensing market News Plaintiffs envision would effectively grant copyright owners veto power over the development of a transformative technology. *See Google*, 593 U.S. at 39 (warning against copyright enforcement that would create "a lock limiting the future creativity of new programs"); *White v. West Publ'g Corp.*, 29 F.Supp.3d 396, 400 (S.D.N.Y. 2014) (finding fair use because "the transactions costs in licensing" the use "would be prohibitively high").

c. Widespread Adoption of the Challenged Use Would Not Cause Cognizable Harm

Unable to show harm from OpenAI's use, News Plaintiffs point instead to Google's AI Overviews—a different product offered by a different company. Although fair use considers whether widespread conduct "of the sort engaged in by the defendant" would cause market harm, *Campbell*, 510 U.S. at 590, it does not permit News Plaintiffs to attribute Google's alleged harms to OpenAI. Google AI Overviews display answers to user queries on Google, which, according to News Plaintiffs, reduces the need of a Google user to click on a link in Google's search

results. NSUF ¶¶ 1782–1786, 1790–1792; MDL ECF 1763 ¶ 95. News Plaintiffs have not established that Google AI Overviews infringe their copyrights, and in fact have not sued Google. Regardless, ChatGPT does not display answers on Google or impact the rate at which Google users click on links shown in Google search results. *See* OSUF ¶¶ 54–56, 240–246, 254. Whatever harm Google allegedly causes, it is not harm from conduct "of the sort engaged in" by OpenAI.

B. News Plaintiffs Are Not Entitled to Summary Judgment on Their Claim of *Prima Facie* Infringement

News Plaintiffs fail to prove as a matter of law that (1) they own each of the asserted articles, (2) the content allegedly copied from them is copyrightable, and (3) the copying was wrongful. *Zalewski v. Cicero Builder Dev., Inc.*, 754 F.3d 95, 100–103 (2d Cir. 2014).

1. The Record Does Not Permit Summary Judgment on the Validity of News Plaintiffs' Copyrights

News Plaintiffs ask the Court to grant them summary judgment on the validity of the copyrights of over 2,000,000 articles, relying primarily on flawed "Rule 1006" summaries. News Plaintiffs have "fail[ed] to show that [they are] entitled to judgment as a matter of law" as to each of these millions of articles. *Vermont Teddy Bear Co. v. 1-800 Beargram Co.*, 373 F.3d 241, 244 (2d Cir. 2004).

a. News Plaintiffs Have Failed to Establish Ownership

News Plaintiffs assert collective work registrations for newspaper issues and compilation registrations for Ziff Davis's website and database registrations. News Mot. at 34. But they seek summary judgment as to individual articles ***within*** those collective works and compilations.[11]

[11] Even as to the copyrighted works the registrations address, there are material disputes of fact that rebut any presumption of validity. *See, e.g.*, RNSUF ¶¶ 335–338 (registration chain of title does not link claimant "Freedom Communications" to named Plaintiff ORB Publishing, LLC); Borck Decl. ¶ 50 (Ziff Davis registered 1,463,999 Asserted Articles in 2025, and 2,581 Asserted

Because News Plaintiffs admittedly do not own the copyright in every article contained in their publications, they must provide undisputed evidence of their ownership, at the time of registration, of each individual article they assert. *See* RNSUF ¶¶ 298, 320–327, 408–409; *Morris v. Bus. Concepts, Inc.*, 283 F.3d 502, 506 (2d Cir. 2002) ("[U]nless the copyright owner of a collective work also owns all the rights in a constituent part, a collective work registration will not extend to a constituent part."); *New York Times Co. v. Tasini*, 533 U.S. 483, 494 (2001) ("Copyright in the collective work ... extends only to the creative material contributed by that author, not to 'the preexisting material employed in the work.'") (quoting 17 U.S.C. § 103(b)).

News Plaintiffs assert a single basis for their alleged ownership of millions of individual articles—that each was authored by an employee or pursuant to a "'work for hire' style agreement[.]" News Mot. at 34–35. The record, however, does not establish who authored each asserted article or the agreements under which the articles were purportedly written. For example, News Plaintiffs rely on article bylines as evidence of an article's writer. Yet their own witnesses admitted that bylines that state "staff" or simply list the newspaper name do not conclusively establish who wrote a given article. *See, e.g.*, RNSUF ¶¶ 306, 372–381, 383.

Similarly, News Plaintiffs rely on employment records to claim that they employed the writers of the articles at the time the articles were written. Yet many of those records do not contain start dates, end dates, or other information establishing employment status. *See, e.g.*, RNSUF ¶ 390. News Plaintiffs' reliance on "Rule 1006" summaries is equally unavailing, as those summaries purport to summarize witness testimony without identifying the alleged "witnesses" or testimony being summarized, *see, e.g.*, Maisel Decl. Ex. A (listing only "Witness"

Articles associated with "BabyCenter" were registered the day Ziff Davis filed suit or after); NSUF ¶¶ 301–304 (The Times registers only its print late edition, yet asserts online versions that may differ).

in "Source" column), and in so doing, invite the Court to engage in impermissible credibility determinations. *Murphy Med. Assocs., LLC v. Cigna Health & Life Ins. Co.*, 2026 WL 1652610, at *5 n.2 (D. Conn. May 8, 2026).

Finally, News Plaintiffs ignore record evidence that precludes summary judgment on validity, such as the Ziff Davis Plaintiffs' registration of material as databases even though, per the testimony of their 30(b)(6) witness, that material does not qualify as a database. *Compare, e.g.*, RNSUF ¶ 399, *with* Compendium of U.S. Copyright Office Practices, 3d ed., § 727.

b. News Plaintiffs Failed to Establish Copyrightability

News Plaintiffs have not provided undisputed evidence showing that each asserted article is sufficiently creative to qualify for copyright protection. *See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991). Many of the asserted articles clearly do not contain the requisite level of creativity. *See, e.g.*, *infra* Section IV.B.2.

2. News Plaintiffs' Purported Evidence of Substantial Similarity Is Insufficient to Grant Summary Judgment for Millions of Works

To establish copyright infringement, a plaintiff must show: (1) actual copying; and then (2) "that the copying amounts to an improper appropriation by demonstrating that substantial similarity to protected material exists between the two works." *Shine v. Childs*, 382 F.Supp.2d 602, 611 (S.D.N.Y. 2005). Substantial similarity is an "extremely close question of fact" that is generally reserved for the factfinder. *Hoehling v. Universal City Studios, Inc.*, 618 F.2d 972, 977 (2d Cir. 1980).

Many of the asserted articles are "manifestly factual" in nature and therefore afforded "at best a 'thin' copyright." *Swatch Grp. Mgmt. Servs. Ltd. v. Bloomberg L.P.*, 861 F.Supp.2d 336, 341 (S.D.N.Y. 2012); *Feist Publ'ns*, 499 U.S. at 348 (The "news of the day ... may not be copyrighted[.]"). For works containing protected and unprotected elements, the substantial

similarity analysis requires comparison of only the protectable elements under a "more discerning" ordinary-observer test. *New York Times Co. v. Microsoft Corp.*, 777 F.Supp.3d 283, 327 (S.D.N.Y. 2025).

News Plaintiffs seek an unprecedented finding of *prima facie* infringement for millions of works based on different expert-created "thresholds" that purportedly show "verbatim overlap" between asserted articles and text purportedly found in training data. News Mot. at 36.[12] Tellingly, each expert applies a different approach. The Times and DNP require at least 80% coverage of 3-word sequences plus a shared 16-word sequence, whereas CIR employs a Jaccard similarity score of at least 0.6. *See id.* at 36–39. Ziff Davis, for its part, relies on either an "Exact Match" (which News Plaintiffs define as including "directional word containment" of at least 95%) or a "Revised Version" (which combines a directional word containment score with a Jaccard similarity threshold). *Id.* at 39–40. In any event, News Plaintiffs cite no authority for any of these approaches, asserting only that courts find infringement when direct evidence shows nontrivial overlap. *Id.* at 35. But none of their cited cases even speak to *prima facie* infringement, much less hold that a plaintiff can establish infringement as a matter of law for millions of works through expert testimony like that at issue here.

Even if the law supported News Plaintiffs' approach, it would still fail because their experts did not distinguish protected expression from unprotected material before calculating whether a work satisfies their chosen "coverage threshold." Filtering out unprotected expression is a necessary first step of the substantial similarity analysis. *See, e.g.*, *Walker v. Time Life Films, Inc.*, 784 F.2d 44, 49–50 (2d Cir. 1986) (excluding similarities of uncopyrightable events that

[12] News Plaintiffs purport to "reserve" the substantial similarity analysis for works falling below their expert-designated threshold scores for trial. News Mot. at 36.

were "reported in the news media"). News Plaintiffs neither address nor apply that rule. Instead, they treat ***any*** overlapping text as *per se* evidence of substantial similarity, even if it comes from a Supreme Court opinion, a third-party newspaper, or a presidential address, or from some other source News Plaintiffs do not own. News Mot. at 36–43.

News Plaintiffs' failure to conduct the qualitative analysis necessary is clear from the works on which they seek summary judgment. For example, they seek summary judgment regarding infringement based on overlapping uncopyrightable content from:

- President Lincoln's Second Inaugural Address
- The Bill of Rights
- A complaint for trademark infringement filed in *Cisco v. Apple*, 3:07-cv-00198 (N.D. Cal. 2007)
- The Supreme Court's opinion in *Planned Parenthood of Southeastern Pennsylvania v. Casey*, 505 U.S. 833 (1992)
- A track listing for a video game soundtrack
- An As Tall As Lions concert schedule

RNSUF ¶¶ 1865–1870.

Additional methodological flaws, detailed in Defendants' *Daubert* motions, independently preclude summary judgment on *prima facie* infringement.[13] The new expert materials submitted at summary judgment have not resolved these methodological flaws and introduce new errors.

[13] Additionally, News Plaintiffs' experts rely extensively on new data, analyses, and opinions prejudicially served after expert discovery closed, and they can be disregarded on that basis. *Cf. Advanced Analytics, Inc. v. Citigroup Glob. Markets, Inc.*, 301 F.R.D. 31, 37–38 (S.D.N.Y.), *objections overruled*, 301 F.R.D. 47 (S.D.N.Y. 2014). The Times and Daily News Plaintiffs rely on expert analyses first disclosed with their summary judgment motion. RNSUF ¶ 1828. Ziff Davis likewise relies on a supplemental expert report served at 2:30 a.m. on the day Plaintiffs filed their motion. RNSUF ¶ 1833. CIR and Intercept rely on a supplemental report served two weeks after the close of expert discovery. *See* RNSUF ¶ 1830; MDL ECF 1452.

NYT & DN Plaintiffs. Dr. Goldstein's totals for purported copies and unique Asserted Works cannot be replicated and match neither News Plaintiffs' brief nor their corrected Rule 56.1 Statement. RNSUF ¶¶ 1826–1827. These discrepancies persist despite substantial revisions for which Dr. Goldstein supplied no corresponding scripts, including a new analysis of the New York Times Annotated Corpus identifying hundreds of thousands of purported matches without explaining how they were determined. RNSUF ¶ 1828. His results also include erroneous "matches" based on overlapping excerpts of third-party books and poetry. RNSUF ¶ 1829.

CIR & The Intercept. CIR seeks summary judgment on 682 works based on Dr. Wenger's analysis. But the unique-work count News Plaintiffs attribute to her report appears nowhere in it, and her analysis includes thousands of works not asserted in News Plaintiffs' Rule 1006 summaries. RNSUF ¶ 1831. Her results also violate her own stated criteria: for example, despite purporting to exclude articles under 200 words, she identifies a 75-word article as an infringing match. RNSUF ¶ 1832.

Ziff Davis. Mr. Delorey's totals for copies and unique Asserted Works attributable to each infringement theory cannot be reproduced from his data and are incorrect based on his own methodology. RNSUF ¶ 1834. His analysis includes thousands of URLs absent from Ziff Davis's Rule 1006 summaries, and his work count differs from the number of works on which Ziff Davis seeks summary judgment. RNSUF ¶¶ 1835–1837. He also reports facially erroneous matches, including, for example, statements by government officials. RNSUF ¶¶ 1838.

C. News Plaintiffs' DMCA Claim Fails as a Matter of Law

Congress enacted the DMCA to protect consumers from misinformation, by prohibiting the (1) intentional (2) removal or alteration of (3) CMI conveyed in connection with a copy of an Asserted Work where (4) the accused knew or had reasonable grounds to know that the removal would induce, enable, facilitate, or conceal copyright infringement. *See Fischer v. Forrest*, 968

F.3d 216, 222–23 (2d Cir. 2020), 17 U.S.C. § 1202(b)(1). News Plaintiffs seek summary judgment on ***millions*** of alleged instances of CMI removal, demanding hundreds of billions of dollars in statutory damages, but offer no evidence that OpenAI did anything beyond extracting useful training data. They move only on the first three of four required elements, "reserving" the second scienter requirement for trial. News Mot. at 75–76. News Plaintiffs' arguments fail as a matter of law, and OpenAI is entitled to summary judgment on their DMCA claims based on undisputed facts. *See* OpenAI-News Mot. at 21–26.

1. News Plaintiffs Cannot Establish Requisite Intent

OpenAI's text extractors convert raw HTML into training text by retaining central human-written content while excluding navigation bars, advertisements, code, and other extraneous material. OpenAI-News Mot. at 21–22. News Plaintiffs concede that this process aimed to maximize human-written text and minimize unuseful data for training, not remove CMI. NSUF ¶¶ 1352–1354; News Mot. at 78.

But "retaining useful portions of information and discarding the rest is not sufficient to demonstrate intent." *Beaulier v. Meta Platforms, Inc.*, 2026 WL 2521848, at *4 (N.D. Cal. Aug. 26, 2026). Mere knowledge that CMI is removed does not satisfy the first scienter requirement. OpenAI-News Mot. at 23. In *McGucken v. Shutterstock, Inc.*, 166 F.4th 361, 370 (2d Cir. 2026), the Second Circuit affirmed summary judgment where CMI removal was incidental to a general metadata-stripping process aimed at other purposes. Here, too, OpenAI processed text to identify readable language useful for pretraining; any CMI removal was likewise incidental to that purpose. News Mot. at 78; OSUF ¶¶ 343–346.

2. News Plaintiffs Cannot Establish CMI Removal

First, the DMCA "requires taking away or changing attribution information from ***an existing work***." *Doe 1 v. GitHub, Inc.*, — F.4th —, 2026 WL 2728464, at *5 (9th Cir. Sept. 16,

2026). But OpenAI's extractors instead transform raw HTML into "materially different and entirely distinct extracted text." OSUF ¶¶ 341–346. They thus create new records "to which plaintiffs' CMI was never attached and from which [OpenAI] could not have 'removed' or 'altered' CMI." *GitHub*, 2026 WL 2728464, at *6.

Second, News Plaintiffs' experts failed to reliably identify CMI removals. They searched only portions of training-data records and treated information as "removed" even when OpenAI indisputably retained it. MDL ECF 2055 at 12 (80% of allegedly removed NYT/DN titles actually retained), 22 (70% of CIR/Intercept titles, 34.8% of author names), 25 (80% of Ziff Davis titles, 94% of publishers). They also counted as "removals" instances where OpenAI's training data came from non-News-Plaintiff sources that never contained News Plaintiffs' CMI in the first place. *Id.* at 21. These error-laden methods do not support summary judgment.

3. News Plaintiffs Have Not Reliably Identified Their Own CMI

Factual disputes over whether News Plaintiffs conveyed CMI with the Asserted Works also preclude summary judgment in their favor. The DMCA defines CMI to include rightsholder-identifying information "conveyed in connection with" the work. 17 U.S.C. § 1202(c).

But, ***first***, ten million instances of News Plaintiffs' "CMI" came from website footers, headers, and navigation bars, ***not*** from the website body containing the asserted work. RNSUF ¶¶ 1839–1840 (4 million NYT/DN footer removals, 6 million Ziff Davis removals attributable to sitewide content). Sitewide content automatically appended to websites and identical across pages is not CMI. *See Pers. Keepsakes, Inc. v. Personalizationmall.com, Inc.*, 975 F.Supp.2d 920, 929 (N.D. Ill. 2013) (copyright notice "at the bottom of every page of the website in the generic website footer" is not CMI "conveyed with" the work). *Pierson v. Infinity Music & Ent., Inc.*, 300 F.Supp.3d 390, 396 (D. Conn. 2018), is inapposite: the problem is the boilerplate's lack of connection to the asserted work, not its distance from it. Indeed, OpenAI's expert found that

over 90% of these "copyright notices" contained registration dates that did not reflect the Asserted Works' publication or registration dates. RNSUF ¶ 1841.

Second, the alleged "CMI" News Plaintiffs' experts have identified is inaccurate and differs from the CMI actually present in their Asserted Works databases in millions of instances. MDL ECF 2055 at 11; RNSUF ¶¶ 1842.

Third, News Plaintiffs' experts count publisher identity as CMI, although it does not inherently qualify under 17 U.S.C. § 1202(c), inflating their tally of alleged "removals" by millions. RNSUF ¶¶ 1843–1844 (7 million NYT/DN publisher "removals" and 3.6 million Ziff Davis publisher "removals").

D. Statutory Damages

1. News Plaintiffs Are Not Entitled to Statutory Damages on a Per-Article Basis

News Plaintiffs assert copyrights registered as compilations but seek statutory damages per article, which would increase maximum damages from billions to hundreds of billions of dollars. News Mot. at 78–80. That request conflicts with the Copyright Act and Second Circuit law: "the Copyright Act limits the copyright holders' statutory damage award to one for each [compilation]." *Bryant v. Media Right Prods., Inc.*, 603 F.3d 135, 141 (2d Cir. 2010); *accord* 17 U.S.C. § 504(c) ("[A]ll the parts of a compilation ... constitute one work.").

News Plaintiffs address neither the statute nor *Bryant*. Instead, they rely on *EMI Christian Music Group v. MP3tunes, LLC*, 844 F.3d 79, 101 (2d Cir. 2016),[14] reasoning that per-song awards were available there because "each individual [song] at issue ... was available for purchase as a single [work]," and so per-article awards are available here. News Mot. at 79. But

[14] News Plaintiffs ask only whether statutory damages, if awarded, should be calculated per article or per compilation. They do not ask this Court to address whether they are, in fact, entitled to statutory damages as to any particular compilation or article.

they identify no evidence that they made available for sale any asserted article individually and simply contend that the articles were "available online for reading and downloading." *Id*. Even under News Plaintiffs' reading of *EMI*, that is not enough.

2. Ziff Davis Is Entitled to No Statutory Damages

Ziff Davis admits it never produced copyright registrations for 180,011 Asserted Works, MDL ECF No. 1826 ("Taylor Decl.") ¶ 28, RNSUF ¶¶ 1864, and identifies no other evidence establishing registration. For the remaining works, Ziff Davis cannot meet its burden of establishing compliance with 17 U.S.C. § 412. *Ushodaya Enters., Ltd. v. V.R.S. Int'l, Inc.*, 64 F.Supp.2d 352, 353 (S.D.N.Y. 1999). Its Rule 1006 summaries list 1,463,999 database registrations purportedly effective in March and April 2025, after the alleged infringement. Borck Decl. ¶ 50.

If, as Ziff Davis suggests, its articles were unpublished, then Section 412 bars statutory damages because the alleged infringement occurred long before registration. *See* Brook Ex. 364, at 4; RNSUF ¶¶ 1845, 1863. If instead its articles were "published" when first posted online, Section 412(2) bars statutory damages because Ziff Davis failed to register them within three months of first publication. *See* Koonce Decl., Ex. A. Either way, Ziff Davis cannot recover statutory damages.

E. News Plaintiffs Are Not Entitled to Summary Judgment on OpenAI's Affirmative Defenses

News Plaintiffs seek summary judgment on multiple OpenAI affirmative defense in a single paragraph. News Mot. at 80. Although a movant may ordinarily satisfy its summary judgment burden by "point[ing] to an absence of evidence on an essential element" for which the nonmovant bears the burden of proof, *see Bustamante v. KIND, LLC*, 100 F.4th 419, 432 (2d Cir. 2024) (cleaned up), News Plaintiffs identify no such "essential element" of any affirmative

defense for which there is an absence of evidence. Their blanket request for summary judgment should be denied.

In any case, OpenAI articulated the basis for its fair use and implied license defenses. OpenAI-News Mot. at 7–21. There is ample record support for OpenAI's other affirmative defenses. For example, there are disputes of fact regarding the timing of News Plaintiffs' knowledge of OpenAI's use of their data to train models and their failure to timely act on purported infringement of their works, such that summary judgment on defenses like waiver, abandonment, forfeiture, laches, failure to mitigate, estoppel, or statute of limitations is inappropriate. *See, e.g.*, RNSUF ¶¶ 1871–1873.

V. CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that the Court deny News Plaintiffs' motion for summary judgment and grant OpenAI's motion for summary judgment.

Dated: October 23, 2026

Respectfully submitted,

LATHAM & WATKINS LLP
Andrew M. Gass (*pro hac vice*)
andrew.gass@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

Elana Nightingale Dawson (*pro hac vice*)
elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)
jgratz@mofo.com
Tiffany Cheung (*pro hac vice*)
tcheung@mofo.com
Caitlin Sinclaire Blythe (*pro hac vice*)
cblythe@mofo.com
425 Market Street
San Francisco, CA 94105
Telephone: 415.268.7000

Lily Li (*pro hac vice*)
YLi@mofo.com
755 Page Mill Road
Palo Alto, CA 94304
Telephone: 650.813.5600

KEKER, VAN NEST & PETERS LLP
Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com
R. James Slaughter (*pro hac vice*)
rslaughter@keker.com
Paven Malhotra
pmalhotra@keker.com
Michelle S. Ybarra (*pro hac vice*)
mybarra@keker.com
Reid P. Mullen (*pro hac vice*)
rmullen@keker.com
Nicholas S. Goldberg (*pro hac vice*)
ngoldberg@keker.com
633 Battery St.
San Francisco, CA 94111
Telephone: 415.391.5400

Attorneys for OpenAI

CERTIFICATE OF COMPLIANCE

In accordance with Local Civil Rule 7.1(c), I hereby certify that the foregoing OPENAI'S OPPOSITION TO NEWS PLAINTIFFS' PARTIAL MOTION FOR SUMMARY JUDGMENT is 8,743 words, exclusive of the caption page, table of contents, table of authorities, and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this memorandum.

Dated: October 23, 2026 By: 