

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

DAILY NEWS, LP; CHICAGO TRIBUNE COMPANY, LLC; ORLANDO SENTINEL COMMUNICATIONS COMPANY, LLC; SUN-SENTINEL COMPANY, LLC; SAN JOSE MERCURY-NEWS, LLC; DP MEDIA NETWORK, LLC; ORB PUBLISHING, LLC; and NORTHWEST PUBLICATIONS, LLC,

Plaintiffs,

v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO, LLC, OPENAI GLOBAL, LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC,

Defendants.

Civil Action No. 1:24-cv-03285-SHS

**PLAINTIFFS' MEMORANDUM OF LAW IN OPPOSITION TO OPENAI
DEFENDANTS' MOTION TO DISMISS**

TABLE OF CONTENTS

I. INTRODUCTION 1

II. BACKGROUND 3

 A. The Publishers’ Investment and Commitment to Local Journalism 3

 B. OpenAI and its Business Model are Built on Mass Copyright Infringement 4

 C. The Publishers File Suit to Protect Their Works, and Defendants File Partial Motions
 To Dismiss 6

III. LEGAL STANDARD 7

IV. ARGUMENT 7

 A. The Publishers’ Direct Infringement Claim is Timely (Count I) 7

 B. The Publishers Have Stated a Contributory Infringement Claim (Count IV) 9

 C. The Publishers Have Properly Pled Their DMCA Claim (Count V) 13

 1. The Publishers Have Standing to Bring Their DMCA Claims 14

 2. The Publishers Have Properly Pled a Section 1202(B)(1) Claim With Respect
 to Training 17

 3. The Publishers Have Properly Pled a Section 1202(B)(3) Claim With Respect
 to Outputs 19

 D. The Copyright Act Does Not Preempt the Publishers’ “Hot-News” Misappropriation
 Claim (Count VI) 20

 E. The Publishers Properly Pled that the Asserted Trademarks are Famous (Count VII) 23

V. CONCLUSION 25

TABLE OF AUTHORITIES

Cases

<i>A&M Recs., Inc. v. Napster, Inc.</i> , 239 F.3d 1004 (9th Cir. 2001)	10, 11
<i>A.V.E.L.A., Inc. v. Est. of Marilyn Monroe, LLC</i> , 131 F. Supp. 3d 196 (S.D.N.Y. 2015).....	24
<i>ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.</i> , 667 F. Supp. 3d 411 (S.D. Tex. 2023)	20
<i>Argo Contracting Corp. v. Paint City Contractors, Inc.</i> , No. 00-CV-3207, 2000 WL 1528215 (S.D.N.Y. Oct. 16, 2000)	12
<i>Arista Records LLC v. Usenet.com</i> , 633 F. Supp. 2d 124 (S.D.N.Y. 2009).....	11, 12, 13
<i>Arista Records, Inc. v. Flea World, Inc.</i> , No. 03-CV-2670, 2006 WL 842883 (D.N.J. Mar. 31, 2006).....	12
<i>Arista Records, Inc. v. Mp3Board, Inc.</i> , No. 00-CV-4660, 2002 WL 1997918 (S.D.N.Y. Aug. 29, 2002).....	9
<i>Arista Records, LLC v. Doe 3</i> , 604 F.3d 110 (2d Cir. 2010)	9
<i>Ashcroft v. Iqbal</i> , 556 U.S. 662 (2009).....	7
<i>Barclays Capital Inc. v. Theflyonthewall.com, Inc.</i> , 650 F.3d 87606 (2d Cir. 2011).....	21
<i>Bell Atl. Corp. v. Twombly</i> , 550 U.S. 544 (2007).....	7
<i>Bose BV v. Zavala</i> , No. 09-CV-11360, 2010 WL 152072 (D. Mass. Jan. 14, 2010)	15
<i>BWP Media USA Inc. v. Hollywood Fan Sites, LLC</i> , 69 F. Supp. 3d 342 (S.D.N.Y. 2014)	10
<i>Capitol Records, LLC v. ReDigi Inc.</i> , 934 F. Supp. 2d 640 (S.D.N.Y. 2013)	9, 10, 13
<i>Car-Freshner Corp. v. Meta Platforms, Inc.</i> , No. 22-CV-1305, 2023 WL 7325109 (N.D.N.Y. 2023)	24
<i>CDC Newburgh Inc. v. STM Bags, LLC</i> , No. 22-CV-1597, 2023 WL 6066136 (S.D.N.Y. Sept. 18, 2023).....	25
<i>Chicago Bd. Options Exch., Inc. v. Int’l Sec. Exch., L.L.C.</i> , 973 N.E.2d 390 (Ill. App. Ct. 1st Dist. 2012).....	21
<i>CoxCom, Inc. v. Chaffee</i> , 536 F.3d 101 (1st Cir. 2008).....	15

DigitAlb, Sh.a v. Setplex, LLC, 284 F. Supp. 3d 547 (S.D.N.Y. 2018) 25

Doe I v. GitHub, Inc., 672 F. Supp. 3d 837 (N.D. Cal. 2023)..... 18

Erickson Beamon Ltd. v. CMG Worldwide, Inc., No. 12-CV-5105,
2014 WL 3950897 (S.D.N.Y. Aug. 13, 2014)..... 23, 24

Fischer v. Forrest, 286 F. Supp. 3d 590 (S.D.N.Y. 2018)..... 20

FurnitureDealer.Net, Inc. v. Amazon.com, No. 18-CV-232,
2022 WL 891473 (D. Minn. Mar. 25, 2022) 15, 16

Global Brand Holdings, LLC v. Church & Dwight Co., No. 17-CV-6571,
2017 WL 6515419 (S.D.N.Y. Dec. 19, 2017) 25

GMA Accessories, Inc. v. Idea Nuova, Inc., 157 F. Supp. 2d 234 (S.D.N.Y. 2000) 24

Goldman v. Belden, 754 F.2d 1059 (2d Cir. 1985)..... 7

Grand v. Schwarz, No. 15-CV-8779, 2016 WL 2733133 (S.D.N.Y. 2016)..... 24

Granite Partners, L.P. v. Bear, Stearns & Co., Inc., 58 F. Supp. 2d 228 (S.D.N.Y. 1999)..... 12

Heller Inc. v. Design Within Reach, Inc., No. 09-CV-1909,
2009 WL 2486054 (S.D.N.Y. 2009)..... 25

Hirsch v. Rehs Galleries, Inc., No. 18-CV-11864, 2020 WL 917213 (S.D.N.Y. Feb. 26, 2020) .. 8

International News Service v. Associated Press, 248 U.S. 215 (1918) 20

Kelly v. Arriba Soft Corp., 77 F. Supp. 2d 1116 (C.D. Cal. 1999)..... 16

Lane Coder Photography, LLC v. Hearst Corp., No. 22-CV-5071,
2023 WL 5836216 (S.D.N.Y. Sept. 8, 2023)..... 9

Lewittes v. Cohen, No. 03-CV-189, 2004 WL 1171261 (S.D.N.Y. 2004) 24

Mango v. BuzzFeed, Inc., 970 F.3d 167 (2d Cir. 2020)..... 14, 18, 19

McGlynn v. Sinovision Inc., No. 28-CV-4826, 2024 WL 643021 (S.D.N.Y. Feb. 15, 2024) 8

Nat’l Basketball Ass’n v. Motorola, Inc., 105 F.3d 841 (2d Cir. 1997)..... 21

Navatar Grp., Inc. v. DealCloud, Inc., No. 21-CV-1255 (SHS),
2023 WL 1797266 (S.D.N.Y. Feb. 7, 2023)..... 7

NYC Triathlon, LLC v. NYC Triathlon Club, Inc., 704 F. Supp. 2d 305 (S.D.N.Y. 2010) 24, 25

Olusola v. Don Coqui Holding Co., LLC, No. 19-CV-6909,
2021 WL 631031 (E.D.N.Y. Feb. 18, 2021)..... 16

Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007) 19

Pilla v. Gilat, No. 19-CV-2255, 2020 WL 1309086 (S.D.N.Y. Mar. 19, 2020) 20

PK Music Performance, Inc. v. Timberlake, No. 16-CV-1215,
2018 WL 4759737 (S.D.N.Y. Sept. 30, 2018)..... 8

Planck LLC v. Particle Media, Inc., No. 20-CV-10959,
2021 WL 5113045 (S.D.N.Y. Nov. 3, 2021)..... 19

Psihoyos v. John Wiley & Sons, Inc., 748 F.3d 120 (2d Cir. 2014)..... 8

Reiffer v. NYC Luxury Limousine Ltd., No. 22-CV-2374,
2023 WL 4029400 (S.D.N.Y. June 15, 2023) 18

Reilly v. Plot Commerce, No. 15-CV-05118,
2016 WL 6837895 (S.D.N.Y. Oct. 31, 2016)..... 15, 19

Rosen v. Amazon.com, Inc., No. 14-CV-2115, 2014 WL 12597073 (C.D. Cal. May 28, 2014).. 12

Snail Games USA Inc. v. Tencent Cloud LLC, No. 22-CV-02009,
2022 WL 3575425 (C.D. Cal. June 6, 2022) 12

Sohm v. Scholastic Inc., 959 F.3d 39 (2d Cir. 2020) 8

State St. Glob. Advisors Tr. Co. v. Visbal, 431 F. Supp. 3d 322 (S.D.N.Y. 2020)..... 9

Steele v. Bongiovi, 784 F. Supp. 2d 94 (D. Mass. 2011) 16

Stevens v. Corelogic, Inc., 899 F.3d 666 (9th Cir. 2018)..... 18

Tiffany (NJ) Inc. v. eBay Inc., 600 F.3d 93 (2d Cir. 2010) 13

TransUnion LLC v. Ramirez, 549 U.S. 413 (2021) 14, 16

Tremblay v. OpenAI, Inc., No. 23-CV-3223, 2024 WL 557720 (ND. Cal. Feb. 12, 2024).... 18, 20

U.S. v. Aina-Marshall, 336 F.3d 167 (2d Cir. 2003) 9

Van Praagh v. Gratton, 993 F. Supp. 2d 293 (E.D.N.Y. 2014) 24

Viacom Int'l, Inc. v. YouTube, Inc., 676 F.3d 19 (2d Cir. 2012)..... 9

Warner Chappell Music, Inc. v. Nealy, No. 22-1078, slip op (S. Ct. May 9, 2024)..... 8

Statutes

17 U.S.C § 1202..... 13, 17, 19

17 U.S.C. § 411(a) 22

Other Authorities

H.R. No. 94-1476..... 21

I. INTRODUCTION

The opening sections of OpenAI’s motion to dismiss read more like a press release than the preface to a legal argument. OpenAI’s comments on Plaintiffs’ (the “Publishers”) claims, including its *ad hominem* salvo,¹ offer little to support an argument that might justify dismissal of any aspect of this lawsuit. Still, a review of OpenAI’s comments provides important context for its overall approach to the issues raised in the motion to dismiss.

OpenAI begins by referring to this case as a “copycat” of the New York Times’ lawsuit against OpenAI. Of course, the law is the law, theft is theft, and every claim in this case rests on one or more of the ways in which OpenAI has stolen content from the Publishers’ newspapers. That many of the Publishers’ claims in this lawsuit overlap with claims asserted by The New York Times shows that OpenAI has stolen content across a wide swathe of the news industry.

OpenAI explains that its motion does not address “the core issue of whether using copyrighted content to train a generative AI model” is lawful, Mem. at 2, and that its motion is instead directed to the Publishers’ “ancillary claims.” But “ancillary” does not mean “unimportant,” and it certainly does not mean “unmeritorious.” One of the claims OpenAI challenges (Count IV) charges OpenAI with contributory copyright infringement—essentially that OpenAI, through its knowing, reckless, and willfully blind refusal to acknowledge that its entire business model rests on the work of others, has aided and abetted theft of the Publishers’ copyrighted works. The remaining claims OpenAI challenges (Counts V-VIII) rest on federal

¹ In the first paragraph of its Introduction, OpenAI accuses the Publishers’ owner of “mak[ing] billions by gutting local newsrooms.” This accusation adds nothing to OpenAI’s legal arguments, and it reflects OpenAI’s ignorance of the Publishers’ businesses. The Publishers’ owners have a well-deserved reputation for achievements in the newspaper industry – for example, purchasing newspapers (the Boston Herald and Orange County Register) out of bankruptcy and turning them into profitable enterprises that continue to serve their communities. In recent years, the Publishers have continued to win prestigious national and local awards (including the Pulitzer Prize) for local news, national news, investigative reporting, and breaking news. Compl. ¶¶ (40- 47).

statutes (the Digital Millennium Copyright Act (“DMCA”), the Lanham Act), New York state law (New York Gen. Bus. Law § 360-1), and New York common law (unfair competition by misappropriation) that are all directed to unscrupulous business dealings. In this sense, “ancillary” should be read to mean “accompanying,” not “subsidiary.”

OpenAI concludes its opening with factual assertions that are (1) wholly unsuited to a motion to dismiss, and (2) either wrong, miss the point, or both. For example, OpenAI says that the Publishers had to “coax” certain outputs from OpenAI’s products in ways that “no normal user would ever attempt.” Wrong on two counts: *First*, OpenAI cherry-picks its examples, and ignores others spelled out in the Complaint in which little prompting was needed. Specifically, the Complaint shows that many of the prompts given to Defendants’ chatbots simply involved asking for a summary of an article’s first few paragraphs, followed by a request for the actual text. *Id.* ¶¶ 98 (“Please tell me about the 2017 Chicago Tribune article entitled ‘What to do with a broken Illinois: Dissolve the Land of Lincoln.’ Please format your response a summary of the facts followed by the actual text.”), 100, 102, 104, 106, 108, 110, 112. Other prompts simply ask for the articles’ text. *Id.* ¶¶ 118 (“Please provide the first seven paragraphs of the Denver Post article entitled ‘A lunar eclipse visits Denver Sunday, but it may not be noticeable.”), 121, 124, 127, 130, 133, 136. *Second*, if a user can ask OpenAI’s products for verbatim recitations of news articles, that shows that those articles reside in OpenAI’s models in violation of the Publishers’ copyrights, *and* that OpenAI’s products are a direct substitute for a local newspaper. OpenAI does not recognize these key points, further highlighting its willful blindness to the nature and consequences of its actions.

Finally, OpenAI argues *ipse dixit* that its “regurgitation” (*i.e.*, reproduction of the Publishers’ copyrighted news content) and “hallucination” (*i.e.*, false attribution of output to one

of the Publishers' papers) is "fringe," "uncommon," and "unintended." How uncommon can regurgitation and hallucination be when this Court now has before it a multitude of examples from *nine different newspapers, including some of the nation's largest*, showing exactly the same types of results from OpenAI's products?

II. BACKGROUND

A. THE PUBLISHERS' INVESTMENT AND COMMITMENT TO LOCAL JOURNALISM

Collectively, the Publishers publish eight local newspapers that cover the news and information in local communities across the United States. Compl. ¶¶ 21-28. These publications, some of which have been in circulation for over 100 years, are long-standing pillars in the communities they serve. *Id.* ¶¶ 41-43, 45-46. Their impact extends beyond these local communities, because these newspapers are available in print, online, and in mobile applications throughout the United States. *Id.* ¶¶ 21-28. For example, local reporting by the Publishers like the *Chicago Tribune's* coverage of harmful pharmacy practices, and the *Daily News's* coverage on the health of 9/11 first responders, has received recognition across the country, including Pulitzer Prizes and coverage from national media outlets. *Id.* ¶¶ 9, 40, 41.

The Publishers have invested billions of dollars in investigating and reporting local news stories. *Id.* ¶ 7. The Publishers sustain this commitment to local journalism through revenues generated from subscriptions, licensing, and advertising. *Id.* ¶ 48. The Publishers protect their investment in local journalism by, *inter alia*, keeping some of their content behind a paywall, registering their copyrights, and appending copyright notices and other copyright management information ("CMI") to articles. *Id.* ¶ 49.

B. OPENAI AND ITS BUSINESS MODEL ARE BUILT ON MASS COPYRIGHT INFRINGEMENT

Defendant OpenAI was formed in 2015, purportedly as a “non-profit artificial intelligence research company.” *Id.* ¶ 52. OpenAI shed its nonprofit status in 2019, and is now a full-blown commercial enterprise, with a market value as high as \$90 billion. *Id.* ¶ 54. From 2018 through 2023, OpenAI developed a series of “large language models” (or “LLMs”) that function by copying millions of works and using them to predict words that are likely to follow a given string of text. *Id.* ¶¶ 73-75. Defendants developed the GPT model in 2018, followed by GPT-2 in 2019, GPT-3 in 2020, GPT-3.5 in 2022, and GPT-4 in 2023. *Id.* ¶ 80.

Although OpenAI avoids using the word “copying,” there is no real dispute that the training process involves copying and storing encoded copies of works in computer memory. *Id.* ¶ 75. Training these models involved collecting and storing text content to create training datasets, and then processing that content through the GPT models. *Id.* ¶ 81. The “knowledge cutoff” date for the GPT models has shifted from as early as September 2021 to as recent as December 2023, suggesting that Defendants continue to scrape content for their training datasets. *Id.* ¶ 89. OpenAI worked with Microsoft to build these training datasets, which OpenAI again does not dispute contain the Publishers’ content. *Id.* ¶¶ 63-71. The exact number of the Publishers’ works that Defendants copied to train their models is currently unknown, most importantly because Defendants have not publicly disclosed the makeup of the datasets used to train GPT-3 and each subsequent model. *Id.* ¶¶ 55, 81.

Before filing suit, the Publishers conducted an investigation into Defendants’ infringement and uncovered overwhelming evidence to support the claims. The Complaint demonstrates that Defendants’ models “memorized” copies of the Publishers’ works. *Id.* ¶¶ 98-113; Ex. J. The GPT-4 model outputs verbatim copies of significant portions of the Publishers’

works and/or detailed summaries of those works when prompted. *Id.* That memorization is a product of how the models were trained. *Id.* ¶ 96. Defendants knew or should have known they were infringing the Publishers’ copyrights, particularly because of the Publishers’ CMI on its content, publications, and websites. *Id.* ¶¶ 157-59. Not only did Defendants ignore this CMI, but they also designed the training process and used web scrapers to remove it. *Id.* ¶¶ 159-161.

Illegal copying to build datasets and train the models is just one aspect of the Publishers’ case. Defendants also commit copyright infringement through their user-facing products, including Copilot (formerly Bing Chat) and ChatGPT. These products were built on and are powered by the infringing models, and they separately violate the Publishers’ copyrights through the outputs they provide in response to user queries. That infringement takes at least two forms: (1) showing copies and/or derivatives of the Publishers’ works that were copied to build the model, and (2) showing synthetic search results that copy and/or paraphrase the Publishers’ works retrieved and copied in response to user search queries in real-time. *Id.* ¶¶ 96-97.

ChatGPT, a text-generating chatbot that mimics natural language in response to user prompts, initially produced only the first type of infringing output. *Id.* ¶ 58. After its November 2022 release, ChatGPT became an instant viral sensation, reaching over 100 million users in three months. *Id.* As shown in numerous examples throughout the Complaint, ChatGPT will display copies or derivatives of the Publishers’ works memorized by the underlying models. *Id.* ¶¶ 98-113. Nevertheless, OpenAI’s terms of use say that end-users “own the Output”. *Id.* ¶ 168. Then, in May 2023, came “Browse with Bing,” a plugin to ChatGPT that uses Microsoft’s Bing search product to access the Internet. This enabled ChatGPT to retrieve content beyond what was included in the underlying model’s training dataset. *Id.* ¶ 114. Despite the widely publicized finding that “Browse with Bing” circumvented paywalls and copied verbatim text from

copyrighted articles, OpenAI later launched its Custom GPT store, offering Custom GPTs that likewise evade paywalls and output the verbatim text of copyright-protected articles. *Id.* ¶¶ 145, 147-151. Such “synthetic search” products combine an LLM’s ability to mimic human expression—including the Publishers’ expression—with the ability to generate natural language copies and/or summaries of search results, including the Publishers’ works. *Id.*

Defendants know their products infringe on the Publishers’ copyrights, and that end-users use their products to do so. *Id.* ¶¶ 142-147. Indeed, Defendants have the ability to monitor and terminate users that infringe the rights of copyright owners such as the Publishers, but they choose not to do so. *Id.* ¶¶ 153-155. Defendants benefited tremendously from training on and using the Publishers’ content without paying. *Id.* ¶¶ 177-183, 185. OpenAI is on pace to generate more than \$4 billion in revenue in 2025, *id.* ¶ 180, and Microsoft’s deployment of the GPT-based models throughout its product line helped boost its market cap by a trillion dollars in the past year alone, *id.* ¶ 16. Despite the Publishers’ efforts to limit free access to their content, Defendants free-ride off the Publishers’ investment, misappropriating a huge amount of the Publishers’ copyrighted content, all without paying fair compensation. *Id.* ¶¶ 185-188. Defendants’ unlawful conduct threatens to divert readers away from the Publishers’ content and websites, threatening the revenue that funds their local reporting. *Id.*

C. THE PUBLISHERS FILE SUIT TO PROTECT THEIR WORKS, AND DEFENDANTS FILE PARTIAL MOTIONS TO DISMISS

On April 30, 2024, the Publishers filed the Complaint against Microsoft and OpenAI, asserting claims for copyright infringement, vicarious copyright infringement, contributory copyright infringement, violations of the DMCA, unfair competition by misappropriation, and trademark dilution under the Lanham Act and New York law. The Publishers seek monetary relief and an injunction to stop Defendants’ unlawful conduct. On June 11, 2024, OpenAI filed

this Motion seeking dismissal of (i) a narrow portion of the copyright infringement claim as untimely; (ii) the contributory infringement claim, (iii) the DMCA claim, (iv) the unfair competition by misappropriation claim, (v) and the trademark dilution claim. The Publishers now submit this Opposition to explain why OpenAI's motion should be denied in full.

III. LEGAL STANDARD

In reviewing a motion pursuant to Fed. R. Civ. P. 12(b)(6), a court must accept as true all of the factual allegations in the complaint, draw inferences from those allegations in the light most favorable to plaintiff, and construe the complaint liberally. *Navatar Grp., Inc. v. DealCloud, Inc.*, No. 21-CV-1255 (SHS), 2023 WL 1797266, at *1 (S.D.N.Y. Feb. 7, 2023). To survive a Rule 12(b)(6) motion, a complaint need only “contain sufficient factual matter, accepted as true, to ‘state a claim to relief that is plausible on its face.’” *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007)). This standard does not call for detailed factual allegations, nor does it impose a probability requirement at the pleading stage. *Navatar Grp.*, 2023 WL 1797266, at *1. Instead, a complaint need only “raise a reasonable expectation that discovery will reveal evidence” of unlawful conduct. *Twombly*, 550 U.S. at 545. The court's function, therefore, is “not to weigh the evidence that might be presented at trial but merely to determine whether the complaint itself is legally sufficient.” *Goldman v. Belden*, 754 F.2d 1059, 1067 (2d Cir. 1985).

IV. ARGUMENT

A. THE PUBLISHERS' DIRECT INFRINGEMENT CLAIM IS TIMELY (COUNT I)

OpenAI's statute-of-limitations argument is narrow, addressing only “OpenAI's creation and use of training datasets for GPT-2 and GPT-3.” Mem. at 9. These arguments do not challenge the Publishers' claims as to the “orders of magnitude more powerful” GPT-3.5 and

GPT-4 models developed in 2022 and 2023, respectively. Compl. ¶¶ 56, 58. Nor do they challenge the recent deployment of Defendants’ user-facing products, e.g., ChatGPT and Copilot. *Id.* ¶ 69. OpenAI’s motion also does not seek dismissal as to Defendants’ work on the GPT-2 and GPT-3 models or the training datasets after April, 30, 2021. *Id.* ¶ 89.

As OpenAI concedes, this Circuit’s “precedent holds that the discovery rule applies in copyright cases.” Mem. at 9 (citing *Psihoyos v. John Wiley & Sons, Inc.*, 748 F.3d 120, 124-25 (2d Cir. 2014)); *Sohm v. Scholastic Inc.*, 959 F.3d 39, 50 (2d Cir. 2020) (claims “do[] not ‘accrue’ until the copyright holder discovers, or with due diligence should have discovered, the infringement”).² OpenAI “bears the burden of proof” on this issue because “the statute of limitations [is] an affirmative defense.” *Hirsch v. Rehs Galleries, Inc.*, No. 18-CV-11864, 2020 WL 917213, at *5 (S.D.N.Y. Feb. 26, 2020). OpenAI must “establish[] that [the Publishers] should be charged with constructive notice of the alleged infringement” before the limitations period. *McGlynn v. Sinovision Inc.*, No. 28-CV-4826, 2024 WL 643021, at *2 (S.D.N.Y. Feb. 15, 2024). That burden is a heavy one, and OpenAI has not shown that it is “clear from the face of the complaint” that the “claims are time-barred”—*i.e.*, that by April 30, 2021, the Publishers should have been aware of the infringement relating to OpenAI’s creation and use of training datasets for GPT-2 and GPT-3. *Id.* At this early juncture, “even some doubt” necessitates denial. *PK Music Performance, Inc. v. Timberlake*, No. 16-CV-1215, 2018 WL 4759737, at *7 (S.D.N.Y. Sept. 30, 2018). OpenAI’s motion on this issue should be denied.

² The Supreme Court’s recent decision in *Warner Chappell Music, Inc. v. Nealy*, No. 22-1078, slip op. at 4 (S. Ct. May 9, 2024), did not overrule that precedent, and contrary to OpenAI’s suggestion, Mem. at 9-10, n. 16, the majority’s opinion expressed no “doubt” as to the applicability of the discovery rule in copyright cases. Given binding Second Circuit precedent on the applicability of the discovery rule in copyright cases, OpenAI’s motion on this issue seems to be intended solely to preserve the issue in case the Supreme Court overrules that still-binding Second Circuit precedent down the road.

B. THE PUBLISHERS HAVE STATED A CONTRIBUTORY INFRINGEMENT CLAIM (COUNT IV)

“Contributory infringement occurs where ‘one . . . with knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another.’” *Capitol Records, LLC v. ReDigi Inc.*, 934 F. Supp. 2d 640, 658 (S.D.N.Y. 2013) (“ReDigi”). OpenAI’s only argument is that the Publishers did not adequately plead OpenAI’s “knowledge” of infringing outputs. “The knowledge standard is an objective one; contributory infringement liability is imposed on persons who ‘*know or have reason to know*’ of the direct infringement.” *Arista Records, LLC v. Doe 3*, 604 F.3d 110, 118 (2d Cir. 2010) (“*Doe 3*”) (emphasis added). A number of cases in this district have confirmed this point.³ The Publishers have alleged OpenAI’s actual and constructive knowledge of end-user infringement in at least five independent and legally sufficient ways, none of which OpenAI directly challenges.

First, “OpenAI’s Custom GPT Store contains numerous Custom GPTs specifically designed to circumvent the Publishers’ paywalls,” including the “Remove Paywall CustomGPT, designed to ‘retrieve websites from RemovePaywall.com and provide the text content to bypass paywalls legally’ and a ‘News Summarizer’ Custom GPT that encourages users to ‘save on subscription costs’ and ‘skip paywalls just using the link text or URL.” Compl. ¶ 147. As their names and descriptions show, these custom GPTs induce, cause, or materially contribute to the

³ *ReDigi*, 934 F. Supp. 2d at 658; *Doe 3*, 604 F.3d at 118; *see also Arista Records, Inc. v. Mp3Board, Inc.*, No. 00-CV-4660, 2002 WL 1997918, at *7 (S.D.N.Y. Aug. 29, 2002) (Stein, J.) (denying summary judgment on contributory copyright infringement claim because there was “evidence from which a jury could find that [defendant] possessed constructive knowledge of infringement”); *State St. Glob. Advisors Tr. Co. v. Visbal*, 431 F. Supp. 3d 322, 358 (S.D.N.Y. 2020) (“knowledge may be actual or constructive”). Willful blindness to copyright infringement is not an excuse, and it is also sufficient to satisfy the knowledge requirement. *E.g., Lane Coder Photography, LLC v. Hearst Corp.*, No. 22-CV-5071, 2023 WL 5836216, at *4 (S.D.N.Y. Sept. 8, 2023) (internal citations omitted). A person is “willfully blind” or engages in “conscious avoidance” amounting to knowledge where the person was aware of a high probability of the fact in dispute and consciously avoided confirming that fact. *Viacom Int’l, Inc. v. YouTube, Inc.*, 676 F.3d 19, 35 (2d Cir. 2012) (quoting *U.S. v. Aina-Marshall*, 336 F.3d 167, 170 (2d Cir. 2003)).

infringing conduct of end-users that bypass the Publishers’ paywalls to obtain unauthorized copies of the Publishers’ works. *See A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1021-22 (9th Cir. 2001) (“*Napster*”), as amended (Apr. 3, 2001), *aff’d sub nom. A&M Records, Inc. v. Napster, Inc.*, 284 F.3d 1091 (9th Cir. 2002) (“sufficient knowledge exists to impose contributory liability” because defendant had “knowledge that *specific* infringing material is available using its system, that it could block access to the system by suppliers of the infringing material, and that it failed to remove the material”). The Complaint further alleges that OpenAI knew or should have known about these features, as shown by OpenAI’s “representation that it ‘set up new systems to help review GPTs against [OpenAI’s] usage policies’ and that it ‘continue[s] to monitor and learn how people use GPTs.’” Compl. ¶ 147; *see BWP Media USA Inc. v. Hollywood Fan Sites, LLC*, 69 F. Supp. 3d 342, 356 (S.D.N.Y. 2014) (finding allegations sufficient to survive dismissal based on Defendants’ alleged monitoring of host websites).

Second, the Complaint alleges that OpenAI’s own public statements confirm that it knew or reasonably should have known about end-user infringement. For example, OpenAI was aware that “users were using ChatGPT’s Browse with Bing plug-in to circumvent paywalls” based on “Defendants’ own acknowledgement of the issue on its website.” Compl. ¶ 145; *see ReDigi*, 934 F. Supp. 2d at 658 (Court had “little difficulty concluding that [defendant] knew or should have known that its service would encourage infringement” because of statements on its own website.). Additionally, in late 2023 (after the release of ChatGPT), OpenAI CEO Sam Altman “clashed with OpenAI board member Helen Toner over a paper that Toner wrote criticizing the company over ‘safety and ethics issues related to the launches of ChatGPT and GPT-4, *including regarding copyright issues.*’” Compl. ¶ 139 (emphasis added). These facts support the Publishers’ allegation that OpenAI knew it was contributing to copyright infringement. *See*

Arista Records LLC v. Usenet.com, 633 F. Supp. 2d 124, 155 (S.D.N.Y. 2009) (“*Usenet*”) (relying on “defendants’ employees’ own statements” to find knowledge requirement met).

Third, the Complaint alleges that “Defendants knew or reasonably should have known that training the GPT models on the Publishers’ Works would result in the GenAI products’ outputting material that infringes the Publishers’ Works” because the “Defendants know that the GPT models have the propensity to ‘memorize’ training materials such that the GPT models regurgitate those training materials in response to prompts.” Compl. ¶ 144; *id.* ¶¶ 83-85. OpenAI admits as much in its Motion by asserting that “[t]raining data regurgitation is a *problem* that researchers at OpenAI and elsewhere work hard to address....” Mem. at 6. OpenAI could not have been “work[ing] hard” to address this “problem” without being aware of it.

Fourth, the Complaint alleges that OpenAI knew or should have known about end-user infringement from widespread reporting on the issue. The Complaint cites “widely publicized reporting that users were using ChatGPT’s Browse with Bing plug-in to circumvent paywalls,” Compl. ¶ 145, and “to create disinformation, misinformation, or simply poor replications of newspapers’ copyrighted content on AI-generated ‘pink-slime’ news sites,” *id.* ¶ 146.

Fifth, the Complaint alleges that “Defendants have the ability to monitor users that infringe the rights of copyright owners such as the Publishers,” *id.* ¶ 153, and “control the output of their GenAI products,” *id.* ¶ 156. Why would OpenAI need to issue terms prohibiting its users from infringing others’ intellectual property, or implement (ineffective) guardrails to prevent the output of copyrighted content, if it did not believe that end-users would seek to use its products to infringe copyrights? *See Napster*, 239 F.3d at 1021 (“[I]f a computer system operator learns of specific infringing material available on his system and fails to purge such material from the system, the operator knows of and contributes to direct infringement.”).

All of these points, individually and especially collectively, many based on independent analysis or OpenAI's own statements, are more than adequate to withstand a motion to dismiss. And, of course, the best evidence of OpenAI's knowledge is in its own possession, rendering a finding about OpenAI's knowledge inappropriate, and best reserved for the summary judgment stage or trial. *See Argo Contracting Corp. v. Paint City Contractors, Inc.*, No. 00-CV-3207, 2000 WL 1528215, at *1 (S.D.N.Y. Oct. 16, 2000) (Stein, J.) (denying motion to dismiss to "allow discovery to take place" where relevant evidence "'may be found within the defendant's possession'" (quoting *Granite Partners, L.P. v. Bear, Stearns & Co., Inc.*, 58 F. Supp. 2d 228, 251 (S.D.N.Y. 1999))); *Snail Games USA Inc. v. Tencent Cloud LLC*, No. 22-CV-02009, 2022 WL 3575425, at *5 (C.D. Cal. June 6, 2022) (knowledge of infringement is best addressed at summary judgment stage); *see also Rosen v. Amazon.com, Inc.*, No. 14-CV-2115, 2014 WL 12597073, at *2 (C.D. Cal. May 28, 2014) ("[T]he question of [defendant's] knowledge, which is key to contributory liability, cannot be answered in the absence of admissible evidence.").

Instead of addressing these allegations, OpenAI argues that the Publishers' contributory infringement claim is somehow limited to "the example outputs" cited in the Complaint. Mem. at 10. OpenAI's argument is based on the incorrect premise that the Publishers were obligated to identify, at the pleading stage, every third party who has infringed specific articles by way of Defendants' products. OpenAI is wrong. *See Usenet*, 633 F. Supp. 2d at 154 ("[K]nowledge of specific infringements is not required to support a finding of contributory infringement."); *Arista Records, Inc. v. Flea World, Inc.*, No. 03-CV-2670, 2006 WL 842883, at *14 (D.N.J. Mar. 31, 2006) ("Defendants are incorrect that Plaintiffs are required to prove that Defendants had knowledge of 'specific infringement(s)' at the time the Defendants materially contributed to the direct infringement."). The Publishers need only allege that OpenAI "knew or should have

known that its service would encourage infringement.” *ReDigi*, 934 F. Supp. 2d at 658; *see also Usenet*, 633 F. Supp. 2d at 155 (knowledge requirement met based on allegations concerning the “widespread availability of copyrighted entertainment media” on “Defendants’ servers”).

OpenAI cites *Tiffany (NJ) Inc. v. eBay Inc.*, 600 F.3d 93 (2d Cir. 2010) (Mem. at 11, n.17), but that case applied the higher standard for contributory trademark infringement claims, which the *Tiffany* court recognized is different from the copyright standard. *Id.* at 108.

Finally, if the Publishers’ pleading that Defendants knew or should have known of the direct infringement by end users is found insufficient (it should not be), the Publishers have also pled willful blindness. To the extent Defendants only have a “general awareness” of the infringing activity, their lack of knowledge of specific instances of infringement is a result of their own conscious decision to turn a blind eye to end-user infringement. As outlined above, the Complaint recounts Defendants’ high probability of awareness of the infringement, *see, e.g.*, Compl. ¶ 214, and Defendants may not skirt liability by consciously ignoring direct infringement. *Cf. Tiffany (NJ) Inc.*, 600 F.3d at 109 (“When it has reason to suspect that users of its service are infringing a protected mark, it may not shield itself from learning of the particular infringing transactions by looking the other way.”).

C. THE PUBLISHERS HAVE PROPERLY PLED THEIR DMCA CLAIM (COUNT V)

The Publishers allege that OpenAI violated the DMCA, specifically 17 U.S.C § 1202(b)(1) and (3), when it removed the Publishers’ copyright management information (“CMI”) during training of the GenAI models, Compl. ¶¶ 159-161, 163, and when generating outputs from the GenAI products, *id.* ¶¶ 164-168. Section 1202(b)(1) proscribes intentional removal or alteration of CMI. Section 1202(b)(3) proscribes distribution of works while knowing that CMI has been removed or altered. Both violations require that Defendants knew or had reasonable

grounds to know that their acts would (1) “conceal the Defendants’ own infringement,” and/or (2) “induce, enable, facilitate, or conceal end-users’ infringement resulting from their operation” of the GenAI products. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 171 (2d Cir. 2020). The Complaint provides detailed allegations on these issues. Compl. ¶¶ 159, 164, 167-169, 222.

By removing the Publishers’ CMI during training and in generating output, Defendants prevent CMI from being “retained within the GPT models and/or displayed when the GenAI products disseminate unauthorized copies of the Publishers’ Works to end-users.” Compl. ¶ 159. “[R]emoving the Publishers’ CMI from the Publishers’ Works and outputting the Publishers’ Works without the CMI wrongfully implie[s] that Defendants had permission to use the Publishers’ Works,” and thus hides the Defendants’ own infringement to the public. *Id.* ¶ 169. Distributing the Publishers’ works without CMI also induces, enables, facilitates or conceals end-user infringement. *Id.* ¶¶ 168-169; *see also id.* ¶¶ 147-151.

OpenAI argues that the DMCA claims should be dismissed for lack of standing and for failure to plead the requisite elements for the so-called “training-based” DMCA claim and the “output-based” DMCA claim. These arguments ignore or misconstrue the allegations in the Complaint, and raise factual disputes that cannot be decided in OpenAI’s favor at this stage.

1. The Publishers Have Standing to Bring Their DMCA Claims.

OpenAI makes separate statutory and constitutional standing arguments, both of which are premised on the failure to allege harm as a result of their violation. Mem. at 13-14. To show standing, the Publishers need only show that they “suffer[ed] concrete harm because of the [Defendants’] violation of federal law.” *TransUnion LLC v. Ramirez*, 549 U.S. 413, 426-27 (2021). The Publishers have alleged ample facts “sufficient to show that a chain of events resulting from the removal of the CMI could result in harm to” the Publishers.

FurnitureDealer.Net, Inc. v. Amazon.com, No. 18-CV-232, 2022 WL 891473, at *26 (D. Minn. Mar. 25, 2022) (“[B]y removing the CMI from FDN's Descriptions and placing the FDN Descriptions on Amazon product detail pages Defendants have significantly increased the likelihood of third-party infringement in addition to the cost of an investigation. This is sufficient to demonstrate constitutional standing.”); *see also CoxCom, Inc. v. Chaffee*, 536 F.3d 101, 108 (1st Cir. 2008) (finding that “the chain of events that can result from the sale of filters in CoxCom's service area clearly encompasses a loss of remuneration to CoxCom”); *Bose BV v. Zavala*, No. 09-CV-11360, 2010 WL 152072, at *2 (D. Mass. Jan. 14, 2010) (finding that “Zavala's sales of Media Centers whose region coding was or could be altered could reasonably have been expected to deprive it ‘of the opportunity to earn profits from the sale of legitimate Bose products to consumers’”).

Moreover, OpenAI is wrong that the Complaint does not allege any “imaginable harm here.” Mem. at 13. By stripping the Publishers’ CMI and distributing the Publishers’ works to the public without the CMI, Defendants have created the false impression that they had permission to copy and distribute the Publishers’ works. Compl. ¶ 169. This false impression “fundamentally undermines the Publishers’ business model, which is critically dependent on subscription revenues to fund journalism, because it results in substitutive products for which Defendants seek to charge their customers for access, siphoning off existing and potential customers through their unlawful and uncompensated use of the Publishers’ own products.” Compl. ¶ 185. The Defendants have also undermined the Publishers’ licensing and subscription arrangements by “providing the Publishers’ Works directly to readers.” *Id.* ¶ 187; *Reilly v. Plot Commerce*, No. 15-CV-05118, 2016 WL 6837895, at *11 (S.D.N.Y. Oct. 31, 2016) (removal of

CMI “made it easier for [] potential infringers” to infringe copyrights). The cases OpenAI relies on are inapposite.⁴

OpenAI’s arguments as to the “training-based” claim rest on a red herring—that the Publishers “claim they were harmed in some unspecified way by OpenAI’s maintenance of allegedly unlawful training data in a private dataset that was never disseminated or otherwise made publicly available.” Mem. at 15. Again, OpenAI is wrong. This case is about Defendants’ “creation and operation of” commercial GenAI products “with the Publishers’ content without permission and without paying for the privilege,” Compl. ¶ 4, followed by the dissemination of that material to end users.

In *TransUnion*, the Supreme Court distinguished between the plaintiffs whose information had been released to third parties (and therefore had standing) and the plaintiffs whose information had not been released (and therefore lacked standing for want of injury). 594 U.S. at 434-35. The Publishers plausibly show that they are in the first group because the Publishers’ content that is missing CMI is not just sitting “in a company database,” but is “retained within the GPT models *and/or displayed* when the GenAI products disseminate unauthorized copies of the Publishers’ Works to end-users.” Compl. ¶ 159 (emphasis added). “Constitutional standing under the DMCA is not a high bar,”—and it is one that the Publishers have met here. *FurnitureDealer.Net, Inc.*, 2022 WL 891473, at *26.⁵

⁴ The court in *Steele v. Bongiovi*, 784 F. Supp. 2d 94, 98 (D. Mass. 2011), dismissed a DMCA claim where the alleged injury was that the defendants’ DMCA violations “caused [plaintiff] to lose” a prior copyright lawsuit because the court’s prior holding “was based on the lack of a ‘substantial similarity.’” The second case, *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999), is a summary judgment opinion that does not even address whether the plaintiff was injured by a DMCA violation. There, the court found—after the development of a full record—that the plaintiff “has not shown users of Defendant’s site were any more likely to infringe his copyrights, any of these users did infringe, or Defendant should reasonably have expected infringement.” *Id.*

⁵ Statutory damages are appropriate even if, after discovery, the Publishers could not quantify the precise harm. *Olusola v. Don Coqui Holding Co., LLC*, No. 19-CV-6909, 2021 WL 631031, at *5 (E.D.N.Y. Feb. 18, 2021).

2. The Publishers Have Properly Pled a Section 1202(b)(1) Claim with Respect to Training.

OpenAI’s challenge to what it refers to as the “training-based” Section 1202(b)(1) claim ignores the express allegations in the Complaint and raises factual disputes that cannot be resolved at this stage. Each argument is addressed in turn below.

First, OpenAI argues that the DMCA claim is time barred to the extent it is based on “the building of training datasets” that occurred more than three years ago. Mot. at 17. Not so under the discovery rule, as discussed in Section IV(A), *supra*. Moreover, the Complaint cites ChatGPT’s new “knowledge cutoff date” of December 2023 to show that Defendants are “continuing to create and use unauthorized copies of the Publishers’ Works contained in the training datasets and elsewhere on the internet.” Compl. ¶ 89. The Publishers have alleged that OpenAI is continuing to build training datasets, and in doing so, uses “content extractors that, by design, remove[] the Publishers’ CMI from the Publishers’ works.” *Id.* ¶ 161.

Second, OpenAI argues that “the Complaint lacks allegations about the inclusion (or exclusion) of Plaintiffs’ CMI” in third-party datasets such as Common Crawl. Mem. at 17. Because Common Crawl is a “copy of the Internet” that includes the Publishers’ works, and because Common Crawl extracts files exactly as they are published, the omission of CMI from OpenAI’s datasets can mean only one thing—OpenAI removed it. Compl. ¶¶ 85, 159-161.

Third, OpenAI argues that the Publishers failed to allege facts that show how OpenAI’s removal of CMI during training could “induce, enable, facilitate, or conceal an infringement,” or how OpenAI could have “reasonable grounds to know” it could, because “the point of CMI” is to provide information to “the public,” and “not to govern purely internal databases.” Mem. at 17-18. OpenAI again misconstrues the claim. What the Complaint alleges is that “removing the Publishers’ CMI from the Publishers’ Works and outputting the Publishers’ Works without the

CMI wrongfully implicat[e] that Defendants had permission to use the Publishers' Works," thereby concealing Defendants' infringement and/or inducing, enabling, concealing, or facilitating end-user infringement. *Id.* ¶ 169. As such, the Publishers have sufficiently alleged OpenAI's culpability based on the concealment of its own infringement and the likely future infringement by end-users. *See Mango*, 970 F.3d at 172 (2d Cir. 2020); *Reiffer v. NYC Luxury Limousine Ltd.*, No. 22-CV-2374, 2023 WL 4029400, at *8 (S.D.N.Y. June 15, 2023) (DMCA claim where defendant concealed its own infringement).

Tremblay v. OpenAI, Inc., No. 23-CV-3223, 2024 WL 557720 (ND. Cal. Feb. 12, 2024) is inapposite. The plaintiffs in *Tremblay* focused their claim on how the removal of CMI during training could induce third parties to infringe. *See Tremblay*, 2024 WL 557720, at *4. Here, the Defendants' removal of CMI during training conceals Defendants' own infringement. As to third-party infringement, the plaintiffs in *Tremblay* "alleged that '**every output** from the OpenAI Language Models is an infringing derivative work' without providing any indication as to what such outputs entail—i.e., whether they are the copyrighted books or copies of the books." *Id.* at *5 (emphasis added). By contrast, the Publishers have identified specific infringing outputs that contain verbatim copies of the Publishers' works. Compl. ¶¶ 98-113.

Doe 1 v. GitHub, Inc., 672 F. Supp. 3d 837 (N.D. Cal. 2023), is instructive on the pleading standard for CMI removal claims with respect to third-party infringement. There, the court found that the plaintiffs had pled: (1) "sufficient facts to support a reasonable inference that Defendants intentionally designed the programs to remove CMI from any licensed code they reproduce as output," and (2) that "Defendants knew or had reasonable grounds to know that removal of CMI carried a substantial risk of inducing infringement." *Id.* at 858; *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 675 (9th Cir. 2018) (Because the "statute is written in the future

tense,” plaintiffs “need not show that any specific infringement has already occurred”). Here, the Publishers have sufficiently pled a risk of third-party infringement, as well as the likelihood of concealment of Defendants’ own infringement. *Mango*, 970 F.3d at 172 (“the statutory language requires constructive knowledge of future *concealment*, not future infringement”).

3. The Publishers Have Properly Pled a Section 1202(b)(3) Claim with Respect to Outputs.

OpenAI challenges the Publishers’ Section 1202(b)(3) claim with respect to outputs stripped of CMI. None of OpenAI’s arguments has merit.

First, OpenAI erroneously argues that the Publishers’ claim fails because the Complaint does not allege that OpenAI “distributed” any outputs. Mot. at 18. That is not correct. The Complaint alleges that ChatGPT provides verbatim copies of the Publishers’ works without CMI in “response to user prompts.” Compl. ¶ 97. In the Internet context, to “distribute” means “transmitting the [work] electronically to the user’s computer.” *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1162 (9th Cir. 2007). *Cf. Reilly*, 2016 WL 6837895, at *11 (S.D.N.Y. Oct. 31, 2016) (Defendant “posted the altered image to [its] Website, thereby distributing a work with the knowledge that its CMI was removed in violation of subsection 1202(b)(3)”).

Second, OpenAI argues that this claim applies only to “entire articles” as opposed to “excerpts.” Mem. at 19. OpenAI cites nothing in the statute for that position. Nor could it—the statute proscribes “intentionally remov[ing] or alter[ing] any copyright management information,” without any limitation confining the claim to “entire” copies. 17 U.S.C. § 1202(b). Courts have found allegations virtually identical to, or even less detailed than, the Publishers’ to be more than adequate at the pleadings stage. *Planck LLC v. Particle Media, Inc.*, No. 20-CV-10959, 2021 WL 5113045, at *6 (S.D.N.Y. Nov. 3, 2021) (finding DMCA claim sufficiently pled where Defendants removed CMI from plaintiff’s news content and “distributed infringing

excerpts of Plaintiff’s stories” with the CMI removed); *ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 430 (S.D. Tex. 2023) (“Based on the plain wording of the statute, the Court is not persuaded that the DMCA includes an ‘identical copy’ requirement.”). OpenAI’s cases are far afield.⁶

Third, OpenAI argues this claim should be dismissed because “there was no CMI to remove from the relevant text” because the outputs cited in the Complaint “feature text from the *middle* of articles.” Mem. at 19. OpenAI cites nothing in the statute or caselaw to support this argument, because it would mean (for example) that the DMCA requirements can be by-passed by the simple expedient of displaying only 90% of a copyrighted work. Moreover, the argument rests on an overly narrow review of the “relevant text.” The relevant text here refers to the Publishers’ works from which CMI has been removed. The Publishers have alleged that they place CMI “on every page of [their] websites,” and OpenAI removes that CMI when it produces output copying those works. Compl. ¶ 140. To the extent OpenAI’s argument is that the outputs are not sufficiently similar to the Publishers’ works, that “argument is better reserved for summary judgment when there is a fully developed record.” *Pilla v. Gilat*, No. 19-CV-2255, 2020 WL 1309086, at *12 (S.D.N.Y. Mar. 19, 2020).

D. THE COPYRIGHT ACT DOES NOT PREEMPT THE PUBLISHERS’ “HOT-NEWS” MISAPPROPRIATION CLAIM (COUNT VI)

The Publishers assert a “hot-news” misappropriation claim based on *International News Service v. Associated Press*, 248 U.S. 215 (1918) (“*INS*”). The Publishers’ claim rests on the fact

⁶ The allegedly infringing material in *Fischer* “bears no resemblance whatsoever” to the allegedly copied material. *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609 (S.D.N.Y. 2018), *aff’d*, 968 F.3d 216 (2d Cir. 2020). And the plaintiffs in *Tremblay* did not “provid[e] any indication as to what [the] outputs entail – i.e., whether they are the copyrighted books or copies of the books.” *Tremblay*, 2024 WL 557720, at *5 (N.D. Cal. Feb. 12, 2024).

that synthetic search products built on GPT LLMs, such as OpenAI’s now-disabled Browse with Bing for ChatGPT, use a process known as “grounding” to provide end-users with time-sensitive factual information that the Publishers gathered, and that was not part of the LLMs’ training set. The content gives no attribution to the Publishers, it competes with the Publishers’ news and information services, and threatens the Publishers’ core businesses. *See* Compl. ¶¶ 114-16, 145. OpenAI’s Custom GPT Store also contains numerous Custom GPTs specifically designed to circumvent the Publishers’ paywalls, such as the “Remove Paywall” Custom GPT, designed to “retrieve webpages from RemovePaywall.com and provide the text content to bypass paywalls legally” and the “News Summarizer” Custom GPT that encourages users to “save on subscription costs” and “skip paywalls just using the link text or URL.” *Id.* ¶ 147.

Congress has explained that such a “misappropriation” claim is not preempted by copyright law because “state law should have the flexibility to afford a remedy (under traditional principles of equity) against a consistent pattern of unauthorized appropriation by a competitor of the facts (*i.e.*, not the literary expression) constituting ‘hot’ news, whether in the traditional mold of [*INS*], or in the newer form of data updates from scientific, business, or financial data bases.” H.R. No. 94-1476 at 132, reprinted in 1976 U.S.C.A.N. at 5748, *quoted in Nat’l Basketball Ass’n v. Motorola, Inc.*, 105 F.3d 841, 850 (2d Cir. 1997) (“*NBA*”); *see also Barclays Capital Inc. v. Theflyonthewall.com, Inc.*, 650 F.3d 876, 905-06 (2d Cir. 2011); *Chicago Bd. Options Exch., Inc. v. Int’l Sec. Exch., L.L.C.*, 973 N.E.2d 390, 402 (Ill. App. Ct. 1st Dist. 2012) (claims “predicated on the unauthorized *use* of the providers’ expertise and goodwill” not preempted). That flexibility is especially important where, as here, Defendants’ GenAI products republish news content that is so time-sensitive (*see* Compl. ¶¶ 127-28, alleging the copying of a news article on

the very same day it was published) it would be impossible for the Publishers to obtain a copyright registration in time to seek an injunction under 17 U.S.C. § 411(a).

As to the elements of the claim, OpenAI is wrong that the Publishers did not sufficiently plead: (1) time-sensitive content, and (2) that OpenAI passed off the Publishers’ “hot-news” content as its own, two elements necessary to show lack of preemption. Mem. at 21-22. With respect to time-sensitive content, OpenAI’s argument wrongly focuses on the *training of its LLM*. See *id.* at 21 (citing Compl. ¶ 89 of the “Factual Allegations” and Exhibit J). The training of OpenAI’s LLM is the factual basis for other claims, but the hot-news” misappropriation claim is based on output generated through a process called “grounding,” which uses recent web-posted content that was *not included in any training sets*. See Compl. ¶ 114. As to OpenAI’s reference to the Complaint’s Exhibit J, that material contains memorization studies conducted *using registered works*, thus showing that *Plaintiff’s registered works were within OpenAI’s training data sets*, and that these works support the Publishers’ copyright infringement claims. The “hot-news” claim is based on Defendants’ misappropriation of breaking news stories, not older materials as set out in Exhibit J. See, e.g., Compl. ¶¶ 118-138 (examples with news that was 0-3 days old). These breaking news stories are particularly valuable, and Plaintiffs have won numerous journalism awards tied to this reporting. *Id.* ¶¶ 43, 45.

With respect to the passing-off element (*i.e.*, that OpenAI has passed off the Publishers’ “hot-news” content as its own), OpenAI argues that the pleadings are inconsistent, citing allegations that OpenAI has attempted to “pass of” the Publishers’ “hot-news” content as its own without attribution, that OpenAI has attributed content to the Publishers, and that OpenAI has attributed false, hallucinated content to the Publishers’ publications. See *id.* These are not inconsistencies. Rather, they reflect the fact that OpenAI’s tools provide different results

(sometimes even with identical inputs), and that all three of these outputs are possible (and documented in the Complaint). That the tools *do not always* output “hot-news” data without attribution does not render the Publishers’ misappropriation claim dismissible. To the contrary, the fact that OpenAI sometimes (indeed, often) passes off the Publishers’ breaking news stories without any attribution, such that users may believe the stories to be OpenAI’s own reporting, requires denial of OpenAI’s motion. *See, e.g.*, Compl. ¶¶ 130-31, 168 (verbatim output that did not contain attribution to the Publishers).

E. THE PUBLISHERS PROPERLY PLED THAT THE ASSERTED TRADEMARKS ARE FAMOUS (COUNT VII)

The Publishers are not required at this stage to “*establish*” that the asserted marks are famous. Mem. at 23. The Publishers need only plead sufficient facts to render it *plausible*. *Erickson Beamon Ltd. v. CMG Worldwide, Inc.*, No. 12-CV-5105, 2014 WL 3950897, at *10 (S.D.N.Y. Aug. 13, 2014). The Complaint pleads each of the required elements for fame, including that the Diluted Trademarks are “widely recognized by the general consuming public of the United States” (¶ 235) and have “achieved household recognition through millions of dollars of advertising and promotion across the United States” (¶ 245).

The Complaint further alleges, *inter alia*, that: (i) each of the branded publications has been circulated for over 100 years (¶¶ 40-42, 45, 238-41); (ii) the Diluted Trademarks are protected by multiple Federal Registrations (*id.* ¶ 234 and Exhibit I); (iii) the publications are circulated throughout all 50 states, (*id.* ¶ 243); (iv) the Publishers’ news stories are featured on major national news outlets and have received unsolicited media attention and praise (*id.* ¶¶ 9 and 240); (v) the Publishers collectively own over 40,000 copyright registrations for works published under the Diluted Trademarks (*id.* ¶¶ 21, 22, 25 and 26); (vi) the branded publications are available daily across multiple platforms (*id.* ¶¶ 21, 22, 25 and 26); (vii) each of the

publications achieved national and international fame from being associated with lead (and often award-winning) reporting of some of the most important events in our nation’s and the world’s history (*id.* ¶¶ 40-45); (viii) the Publishers spend hundreds of millions of dollars to operate their publications (*id.* ¶¶ 184, 227); (ix) each of the publications has been widely recognized for its achievements, including the most prestigious (and highly publicized) journalism award, the Pulitzer Prize, including 11 for The Daily News (*id.* ¶ 238), 28 for The Chicago Tribune (*id.* ¶ 239), 9 for The Denver Post (*id.* ¶ 242), and 2 for The Mercury News (*id.* ¶ 42); (x) the publications have widespread circulation across a general audience in the United States (*id.* ¶¶ 238-241); and (xi) millions of consumers access Plaintiffs’ publications circulated under each of the Diluted Trademarks (*id.* ¶ 242).

Collectively these facts sufficiently allege that the Diluted Trademarks are famous. *Car-Freshner Corp. v. Meta Platforms, Inc.*, No. 22-CV-1305, 2023 WL 7325109, at *23–26 (N.D.N.Y. 2023) (finding “***in combination***, and accepting the allegations as true, Plaintiffs have sufficiently stated a dilution claim”) (emphasis added); *A.V.E.L.A., Inc. v. Est. of Marilyn Monroe, LLC*, 131 F. Supp. 3d 196, 216 (S.D.N.Y. 2015) (“***Taken together***, these allegations sufficiently allege that the MONROE Marks are famous”) (emphasis added). Moreover, courts in this circuit that have been tasked with evaluating the adequacy of a plaintiff’s dilution claim under Rule 12, have repeatedly accepted much less than what has been pled here.⁷ By contrast, each of the four cases relied upon by OpenAI is easily distinguishable. In *Heller*, the plaintiff had

⁷ See, e.g., *A.V.E.L.A., Inc.*, 131 F. Supp. 3d at 216; *Van Praagh v. Gratton*, 993 F. Supp. 2d 293, 304–05 (E.D.N.Y. 2014); *Grand v. Schwarz*, No. 15-CV-8779, 2016 WL 2733133, at *5 (S.D.N.Y. 2016); *GMA Accessories, Inc. v. Idea Nuova, Inc.*, 157 F. Supp. 2d 234, 241 (S.D.N.Y. 2000); *Car-Freshner Corp.*, 2023 WL 7325109 at *23–26; *NYC Triathlon, LLC v. NYC Triathlon Club, Inc.*, 704 F. Supp. 2d 305, 321–22 (S.D.N.Y. 2010); *Lewittes v. Cohen*, No. 03-CV-189, 2004 WL 1171261, at *6 (S.D.N.Y. 2004); *Erickson Beamon Ltd. v. CMG Worldwide, Inc.*, No. 12-CV-1505, 2014 WL 3950897, at *10 (S.D.N.Y. Aug. 13, 2014).

pled that its mark was known only within the limited market of furniture buyers and by the general public. *Heller Inc. v. Design Within Reach, Inc.*, No. 09-CV-1909, 2009 WL 2486054, at *4 (S.D.N.Y. 2009) (“trademark is well known to the ‘relevant public interested in contemporary furniture.’”). In *CDC Newburgh* and *DigitAlb*, the plaintiff pled just two or three conclusory statements with no underlying factual allegations. *CDC Newburgh Inc. v. STM Bags, LLC*, No. 22-CV-1597, 2023 WL 6066136, at *18 (S.D.N.Y. Sept. 18, 2023); *DigitAlb, Sh.a v. Setplex, LLC*, 284 F. Supp. 3d 547, 558 (S.D.N.Y. 2018). In *Global Brand Holdings*, the court found that the fame allegations did “not provide any factual detail.” *Global Brand Holdings, LLC v. Church & Dwight Co.*, No. 17-CV-6571, 2017 WL 6515419, at *5 (S.D.N.Y. Dec. 19, 2017).

OpenAI also makes the flawed argument that a publication that has some characteristics of being a “local newspaper” cannot also be nationally recognized or famous. There is no such rule. Nothing prohibits a finding that an entity can be both “local” and still enjoy nationwide recognition. *See, e.g., NYC Triathlon, LLC*, 704 F. Supp. 2d at 321–22 (locally run triathlon race could be famous for dilution purposes). For example, while the San Jose Mercury News may focus on content arising from the local Silicon Valley tech companies, it is the newspaper of record for issues involving companies in Silicon Valley—companies whose stock is owned by, or products used by, virtually every person in the United States.

V. CONCLUSION

For the reasons set forth above, the Publishers respectfully request that the Court deny the motion in its entirety, or in the alternative, grant leave to amend any claim that is dismissed.

Dated: June 25, 2024

/s/ Steven Lieberman

Steven Lieberman (SL8687)
 Jennifer B. Maisel (5096995)
 Robert Parker (*pro hac vice*)
 Jenny L. Colgate (*pro hac vice*)

Mark Rawls (*pro hac vice*)
Kristen J. Logan (*pro hac vice*)
Bryan B. Thompson (6004147)
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
sliberman@rothwellfigg.com
jmaisel@rothwellfigg.com
rparker@rothwellfigg.com
jcolgate@rothwellfigg.com
klogan@rothwellfigg.com
bthompson@rothwellfigg.com

Jeffrey A. Lindenbaum (JL1971)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
The Holyoke-Manhattan Building
80 South Highland Avenue
Ossining, New York 10562
Telephone: (202) 783-6040
Facsimile: (202) 783-6031
jlindenbaum@rothwellfigg.com

Attorneys for Plaintiffs