

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

THE NEW YORK TIMES COMPANY,

Plaintiff,

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. 1:23-cv-11195-SHS

**PLAINTIFF'S MEMORANDUM OF LAW IN OPPOSITION TO MICROSOFT
CORPORATION'S PARTIAL MOTION TO DISMISS (DKT. 64)**

TABLE OF CONTENTS

INTRODUCTION1

BACKGROUND3

 I. The New York Times and its business model are built on world-class journalism.3

 II. Microsoft develops a lucrative generative artificial intelligence business through mass copyright infringement.3

 III. The Times sues to protect its works, and Defendants file partial motions to dismiss.6

LEGAL STANDARD6

ARGUMENT7

 I. The Times states a contributory infringement claim.....7

 A. The Times alleges infringement by end users of Defendants’ products.....7

 B. The Times alleges Microsoft’s knowledge of infringement.9

 II. The Times states DMCA claims on multiple grounds.11

 A. The Times alleges that Microsoft knew or had “reasonable grounds to know” that removal of CMI would conceal or facilitate copyright infringement.....12

 1. Microsoft’s knowledge of the effect of removing CMI during model training.....13

 2. Microsoft’s knowledge of the effect of removing CMI from generative outputs14

 B. The Times alleges removal of CMI from copies of Times works.....16

 C. There is no “judicial consensus” that DMCA claims cannot be brought against GenAI developers.....19

 III. The Copyright Act does not preempt The Times’s “hot-news” unfair competition by misappropriation claim.20

 A. The Times states a non-preempted *INS*-type claim with respect to news content.22

| | | |
|-----------------|---|----|
| B. | The Times states a non-preempted <i>INS</i> -type claim with respect to its recommendations. | 24 |
| CONCLUSION..... | | 25 |

TABLE OF AUTHORITIES

| | Page(s) |
|--|----------------|
| Cases | |
| <i>A&M Recs. Inc. v. Napster Inc.</i> , 114 F. Supp. 2d 896 (N.D. Cal. 2000) | 11 |
| <i>A&M Recs., Inc. v. Napster, Inc.</i> , 239 F.3d 1004 (9th Cir. 2001) | 1, 11 |
| <i>A’Lor Int’l, Ltd. v. Tappers Fine Jewelry</i> , 2012 WL 12921035 (C.D. Cal. Aug. 8, 2012)..... | 18 |
| <i>Adina’s Jewels, Inc. v. Shashi, Inc.</i> , 442 F. Supp. 3d 766 (S.D.N.Y. 2020)..... | 25 |
| <i>Agence France Presse v. Morel</i> , 769 F. Supp. 2d 295 (S.D.N.Y. 2011)..... | 12 |
| <i>Andersen v. Stability AI Ltd.</i> , 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023)..... | 19 |
| <i>Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023)..... | 1 |
| <i>Argo Contracting Corp. v. Paint City Contractors, Inc.</i> , 2000 WL 1528215 (S.D.N.Y. Oct. 16, 2000)..... | 13 |
| <i>Arista Recs., Inc. v. Mp3Board, Inc.</i> , 2002 WL 1997918 (S.D.N.Y. Aug. 29, 2002)..... | 9 |
| <i>Arista Recs., LLC v. Doe 3</i> , 604 F.3d 110 (2d Cir. 2010)..... | 7, 9 |
| <i>Arista Recs. LLC v. Usenet.com, Inc.</i> , 633 F. Supp. 2d 124 (S.D.N.Y. 2009)..... | 9, 10, 11 |
| <i>BanxCorp v. Costco Wholesale Corp.</i> , 723 F. Supp. 2d 596 (S.D.N.Y. 2010)..... | 6, 22, 23 |
| <i>Barclays Cap. Inc. v. Theflyonthewall.com, Inc.</i> , 650 F.3d 876 (2d Cir. 2011)..... | passim |
| <i>Capitol Recs., LLC v. ReDigi Inc.</i> , 934 F. Supp. 2d 640 (S.D.N.Y. 2013)..... | 7, 10 |

| | |
|--|--------|
| <i>DBW Partners, LLC v. Mkt. Sec., LLC</i> , 2023 WL 2610498 (D.D.C. Mar. 23, 2023)..... | 24, 25 |
| <i>Design Basics, LLC v. WK Olson Architects, Inc.</i> , 2019 WL 527535 (N.D. Ill. Feb. 11, 2019) | 18 |
| <i>Devocean Jewelry LLC v. Associated Newspapers Ltd.</i> , 2016 WL 6135662 (S.D.N.Y. Oct. 19, 2016)..... | 12 |
| <i>DiBlasio v. Novello</i> , 344 F.3d 292 (2d Cir. 2003)..... | 23 |
| <i>Doe 1 v. GitHub, Inc.</i> , 2024 WL 235217 (N.D. Cal. Jan. 22, 2024)..... | 19 |
| <i>Doe 1 v. GitHub, Inc.</i> , 672 F. Supp. 3d 837 (N.D. Cal. 2023) | 19 |
| <i>Dolls Kill, Inc. v. Zoetop Bus. Co.</i> , 2022 WL 16961477 (C.D. Cal. Aug. 25, 2022)..... | 18 |
| <i>Falkner v. Gen. Motors LLC</i> , 393 F. Supp. 3d 927 (C.D. Cal. 2018) | 18 |
| <i>Fashion Nova, LLC v. Blush Mark, Inc.</i> , 2023 WL 4307646 (C.D. Cal. June 30, 2023) | 13 |
| <i>Faulkner Press, L.L.C. v. Class Notes, L.L.C.</i> , 756 F. Supp. 2d 1352 (N.D. Fla. 2010)..... | 18 |
| <i>Fin. Info., Inc. v. Moody's Invs. Serv., Inc.</i> , 808 F.2d 204 (2d Cir. 1986)..... | 23 |
| <i>Fischer v. Forrest</i> , 286 F. Supp. 3d 590 (S.D.N.Y. 2018)..... | 18 |
| <i>Frost-Tsuji Architects v. Highway Inn, Inc.</i> , 2015 WL 263556 (D. Haw. Jan. 21, 2015)..... | 18 |
| <i>Gershwin Publ'g Corp. v. Columbia Artists Mgmt.</i> , 443 F.2d 1159 (2d Cir. 1971)..... | 9 |
| <i>Gogo Apparel, Inc. v. True Destiny, LLC</i> , 2020 WL 5578336 (S.D.N.Y. Sept. 17, 2020)..... | 25 |
| <i>Granite Partners v. Bear, Stearns & Co.</i> , 58 F. Supp. 2d 228 (S.D.N.Y. 1999)..... | 13 |

| | |
|--|----------------|
| <i>Greer v. Fox News Media</i> , 2023 WL 2671796 (2d Cir. Mar. 29, 2023)..... | 23 |
| <i>Hirsch v. CBS Broad. Inc.</i> , 2017 WL 3393845 (S.D.N.Y. Aug. 4, 2017)..... | 12 |
| <i>In re DDAVP Antitrust Litig.</i> , 585 F.3d 677 (2d Cir. 2009)..... | 12 |
| <i>Int’l News Serv. v. Associated Press</i> , 248 U.S. 215 (1918)..... | 20 |
| <i>Kelly v. Arriba Soft Corp.</i> , 77 F. Supp. 2d 1116 (C.D. Cal. 1999) | 18 |
| <i>Mango v. BuzzFeed, Inc.</i> , 970 F.3d 167 (2d Cir. 2020)..... | 12, 13, 15, 16 |
| <i>Matthew Bender & Co. v. West Publ’g Co.</i> , 158 F.3d 693 (2d Cir. 1998)..... | 8 |
| <i>Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.</i> , 545 U.S. 913 (2005)..... | 11 |
| <i>ML Genius Holdings LLC v. Google LLC</i> , 2022 WL 710744 (2d Cir. Mar. 10, 2022)..... | 24 |
| <i>Nat’l Basketball Ass’n v. Motorola, Inc.</i> , 105 F.3d 841 (2d. Cir. 1997)..... | 20, 21, 24, 25 |
| <i>Orange Cnty. Choppers, Inc. v. Olaes Enterps., Inc.</i> , 497 F. Supp. 2d 541 (S.D.N.Y. 2007)..... | 25 |
| <i>Planck LLC v. Particle Media, Inc.</i> , 2021 WL 5113045 (S.D.N.Y. Nov. 3, 2021)..... | 17 |
| <i>Shihab v. Complex Media, Inc.</i> , 2022 WL 3544149 (S.D.N.Y. Aug. 17, 2022)..... | 14, 15 |
| <i>Snail Games USA Inc. v. Tencent Cloud LLC</i> , 2022 WL 3575425 (C.D. Cal. June 6, 2022) | 10 |
| <i>Sony Corp. of Am. v. Universal City Studios, Inc.</i> , 464 U.S. 417 (1984)..... | 1, 2, 10, 11 |
| <i>Stevens v. Corelogic, Inc.</i> , 899 F.3d 666 (9th Cir. 2018) | 16 |

Tremblay v. OpenAI, Inc.,
2024 WL 557720 (N.D. Cal. Feb. 12, 2024)9, 14, 19

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

Statutes

17 U.S.C. § 1202(b)15, 16, 17, 19

Other Authorities

H.R. No. 94-1476, reprinted in 1976 U.S.C.C.A.N.20

Kevin Hurler, *OpenAI Pauses ChatGPT’s ‘Browse With Bing’ as Users Bypass Paywalls*, GIZMODO (July 5, 2023)7

Rose Maura Lorre, *The Best Amazon Prime Day 2021 Deals Still Available on Day 2*, WIRECUTTER (June 22, 2021)24

S. Rep. No. 105-190 (1998)11

Tim Heffernan & Ellen Airhart, *How Can I Clear My Home of Wildfire Smoke*, WIRECUTTER (June 7, 2023)24

Yona TR Golding, *The News Media and AI: A New Front in Copyright Law*, COLUMBIA JOURNALISM REV. (Oct. 18, 2023)8

Zainab Choudhry & Aarian Marshall, *News Publishers Are Wary of the Bing Chatbot’s Media Diet*, WIRED (Feb. 11, 2023)8

INTRODUCTION

Microsoft begins its partial Motion to Dismiss (the “Motion”) by comparing Defendants’ generative AI products to VCRs in an effort to invoke *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984), a case affirming a trial verdict that Sony was not liable for contributory infringement for selling VCRs that allowed users to “time-shift” when they watched television shows and movies. Mot. at 1. But Defendants’ generative AI (“GenAI”) models are nothing like VCRs. Sony didn’t copy movies and television shows to build VCRs; Defendants built their AI models by copying millions of Times articles and other copyrighted works without permission or payment. A VCR is not preloaded with millions of programs that can be played back if prompted with an opening scene, *see Sony*, 464 U.S. at 436 (“Petitioners in the instant case do not supply Betamax consumers with respondents’ works; respondents do.”); Defendants’ AI models can display works that were copied to train them with the right prompt. Sony’s VCRs did not generate programming that competed with original works; Defendants are using their AI models to copy and summarize even breaking news articles that users would otherwise seek on a publisher’s website. If VCRs had been built with movies to make movies that compete with movies, or if Sony oversaw the VCR’s infringing users, *Sony* would have gone the other way. *See Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 537–38 (2023) (holding the unlicensed, commercial use of a copyrighted work for the same purpose as the original was infringing); *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1024 (9th Cir. 2001) (finding Napster likely liable for secondary infringement where it “had the right and ability to police its system”).

Microsoft also accuses The Times of opposing the progress of technology, likening this lawsuit to “an all-out effort” by the entertainment industry to stop the VCR. Mot. at 1. But it is entirely possible to comply with copyright law and build GenAI products in a way that promotes

a healthy news ecosystem; Microsoft and its close collaborator OpenAI simply chose not to do so. Instead, Microsoft—the richest company in the world—claims that it was entitled to take what it wanted for free. Calling attention to the market effects of this unlawful conduct is not “doomsday futurology” (Mot. at 3). To the contrary, enforcing copyright protections here will promote a future where both GenAI developers and publishers can flourish.

On the merits, Microsoft’s Motion largely duplicates the arguments made by OpenAI, asking the Court to construe disputed issues of fact in its favor to dismiss The Times’s claims for contributory infringement, Digital Millennium Copyright Act (“DMCA”) violations, and unfair competition by misappropriation. These arguments each fail. *First*, Microsoft’s contributory infringement argument would require the Court to make improper factual findings that Microsoft lacked “awareness” of infringement and that “real-world people” do not “actually use the GPT-based tools” to infringe copyrights, despite The Times’s allegations to the contrary. Mot. at 2, 13; *see* Compl. ¶ 126 (alleging “widely publicized” reports of users’ copyright infringements). *Second*, Microsoft’s DMCA argument similarly asks the Court to trust its word about how “real-world” people use its products and whether Microsoft “should have known” its conduct would conceal or facilitate its infringement. Mot. at 16. *Third*, Microsoft’s misappropriation argument is based on unfounded assertions about whether its products “compete with Times content,” including with regard to its time-sensitive breaking news and product recommendations. Mot. at 22.

None of these issues can be resolved in Microsoft’s favor on a motion to dismiss. Unlike *Sony*—which was decided after a five-week trial—this case is at the pleading stage. Microsoft largely tries to circumvent this reality by citing summary judgment and post-trial decisions as if The Times is required to prove its allegations before the parties have taken any discovery. It is not. This Motion is premature and should be denied in full.

BACKGROUND

I. The New York Times and its business model are built on world-class journalism.

The Times publishes digital and print products, including its core news product, The New York Times, and other publications such as The Athletic, Cooking, Games, and Wirecutter. Compl. ¶ 14. Producing original independent journalism is at the heart of its mission, and The Times invests an enormous amount of time, money, expertise, and talent to do so, covering topics from news to opinion, culture to business, cooking to games, and shopping recommendations to sports. *Id.* ¶¶ 26–27, 32. The Times protects this valuable content by, *inter alia*, keeping much of it behind a paywall, appending copyright notices, links to its terms of service, and other copyright management information (“CMI”) to it, and registering its copyrights. *Id.* ¶¶ 40–41, 49, 124–26, 182. The Times owns more than 3 million registered, copyrighted works. *Id.* ¶ 14.

The Times funds its journalism through revenues derived from subscriptions, advertising, licensing, and affiliate referrals. *Id.* ¶ 40. Maintaining a direct relationship with its audience, as well as traffic to its platforms, is critical to sustaining these revenue streams. *Id.* ¶¶ 41, 52. To facilitate that traffic, The Times allows search engines like Microsoft Bing to index its content, but inherent in this exchange is the premise that search engines will send users directly to The Times’s own websites and mobile applications, rather than exploit The Times’s content to keep users within their own platforms. *Id.* ¶ 52. Defendants’ GenAI products upend this model. *Id.* ¶ 157.

II. Microsoft develops a lucrative generative artificial intelligence business through mass copyright infringement.

Microsoft is the richest company in the world, with a market capitalization of \$3 trillion. Approximately \$1 trillion of that value is derived from GenAI, primarily due to its partnership with OpenAI. Compl. ¶¶ 6, 16, 151. Under the terms of that partnership, Microsoft is currently entitled to 75% of OpenAI’s profits and will own 49% of OpenAI once its initial \$13 billion

investment is repaid. *Id.* ¶ 15. Contrary to Microsoft’s implication that they partnered with OpenAI for the good of humanity (Mot. at 2), the reality is that Defendants’ partnership is expected to be a highly lucrative commercial arrangement for both partners—especially since they are not paying for the content they took from creators to build and operate their products. *Id.* ¶¶ 6, 144, 151–53.

As Microsoft acknowledges (Mot. at 5), Microsoft has been, and continues to be, intimately involved in the training, development, and commercialization of Defendants’ large-language models (“LLMs”). *Id.* ¶ 66. Microsoft and OpenAI “collaborated” to “select[] the training datasets” for their LLMs, including The Times works that comprise those sets. *Id.* ¶ 93. LLMs are algorithms that function by copying millions of works and using them to predict words that are likely to follow a given string of text. *Id.* ¶ 75. There is no dispute that this process involved copying and storing encoded copies of works in computer memory, including Times works. *Id.* ¶ 78.

Before filing suit, The Times sought to determine whether Defendants infringed its copyrights. It uncovered overwhelming evidence that they had. The Complaint includes more than a hundred examples demonstrating that Defendants’ models “memorized” copies of Times works. *Id.* ¶¶ 98–107; Ex. J. The most updated model (GPT-4 LLM) outputs verbatim copies of significant portions of Times works and/or detailed summaries when prompted with the first few words or sentences of articles. *Id.* That memorization is a product of how the models were trained. *Id.* ¶¶ 98, 103. Defendants knew or should have known they were infringing The Times’s copyrights, particularly because The Times places its CMI on all content. *Id.* ¶¶ 124–26. Not only did Defendants ignore this CMI, but they also designed the training process to remove it. *Id.* ¶ 187.

Illegal copying to build datasets and train the models is just one aspect of The Times’s case. Defendants also commit copyright infringement through their user-facing products, including Copilot. These products were built on and are powered by the infringing models, and they

separately violate The Times’s 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 Times works that were copied to build the model, and (2) showing synthetic search results that are substantially similar to Times works retrieved and copied in response to user queries. *Id.* ¶ 102.

Along with OpenAI, Microsoft has developed and commercialized these user-facing products. The first one, ChatGPT, was released in November 2022, and is a text-generating chatbot that mimics natural language in response to user prompts. *Id.* ¶ 61. Then came Bing Chat in February 2023 (currently known as Copilot), a generative AI chatbot feature on Microsoft’s browser’s search engine, which is powered by Defendants’ latest GPT-4 model. *Id.* ¶ 72. Next came “Browse with Bing,” released in May 2023, a plugin to ChatGPT that uses Microsoft’s Bing search engine to access the Internet. This plugin enables ChatGPT to retrieve content beyond what was included in the underlying model’s training dataset. *Id.* ¶¶ 72, 108, 118. “Synthetic search” products like Copilot and Browse with Bing combine an LLM’s ability to mimic human expression—including Times expression—with the ability to generate natural language summaries of search results, including Times works. *Id.* Microsoft’s Copilot and ChatGPT’s “Browse with Bing” feature also allow users to access key portions of Times articles when a user requests them. *Id.* ¶¶ 112, 115, 118, 121, 131 (showing verbatim copying in response to prompts asking for “the first paragraph” of a Times article, followed by a request for the “next paragraph”).

Defendants decided to train their models on protected Times content and then, with those models, developed products that output protected content. *Id.* ¶ 124. After these products were released, The Times expressly put Defendants on notice of their infringement. Compl. ¶¶ 54, 126. Defendants were also aware of copyright infringement after ChatGPT, Browse with Bing, and Bing Chat were released, including widely publicized examples. Compl. ¶ 126.

Meanwhile, The Times has been harmed by Defendants’ free-riding. Defendants have misappropriated almost a century’s worth of copyrighted content without paying fair compensation. *Id.* ¶ 144. Defendants’ unlawful conduct threatens to divert readers from accessing Times content directly from The Times, thus reducing the sources of revenue that The Times uses to fund its journalism. *Id.* ¶ 157.

III. The Times sues to protect its works, and Defendants file partial motions to dismiss.

On December 27, 2023, The Times filed the Complaint against Microsoft and OpenAI, asserting claims for copyright infringement, vicarious copyright infringement, contributory copyright infringement, violations of the Digital Millennium Copyright Act (“DMCA”), unfair competition by misappropriation, and trademark dilution. In addition to monetary relief, The Times seeks an injunction to stop Defendants’ unlawful conduct. On March 4, 2024, Microsoft filed this partial motion to dismiss, seeking dismissal of (i) the contributory infringement claim, (ii) the DMCA claim, and (iii) the unfair competition by misappropriation claim, all of which were also the subject of OpenAI’s Motion to Dismiss. Dkts. 51–52, 64–65. It is unclear why Defendants filed two separate motions to address the same three claims, particularly in light of their close business relationship, rather than coordinating briefing to avoid duplication. In any event, having first responded to OpenAI’s arguments, Dkt. 73, The Times now submits this Opposition to explain why Microsoft’s motion should be denied in full.

LEGAL STANDARD

To overcome a Rule 12(b)(6) motion to dismiss, the Complaint need only “allege enough facts to state a claim to relief that is plausible on its face.” *BanxCorp v. Costco Wholesale Corp.*, 723 F. Supp. 2d 596, 601 (S.D.N.Y. 2010). “[T]he court must accept a plaintiff’s factual allegations as true and draw all reasonable inferences in the plaintiff’s favor.” *Id.* at 600.

ARGUMENT

I. The Times states a contributory infringement claim.

Like OpenAI, Microsoft does not challenge The Times’s direct infringement claim, but instead seeks dismissal of The Times’s alternative contributory infringement claim. The contributory infringement doctrine permits courts “to impose secondary liability where ‘just’ and appropriate.” *Capitol Recs., LLC v. ReDigi Inc.*, 934 F. Supp. 2d 640, 658 (S.D.N.Y. 2013), *aff’d*, 910 F.3d 649 (2d Cir. 2018). It “occurs where ‘one . . . with knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another.’” *Id.* (citing *Arista Recs., LLC v. Doe 3*, 604 F.3d 110, 118 (2d Cir. 2010)).

Relying almost exclusively on cases decided at summary judgment or trial (six of the eight cases cited were decided after the pleading stage), Microsoft argues that The Times does not allege (i) infringing conduct by end users, and (ii) Microsoft’s knowledge of such conduct. Both arguments ignore the Complaint’s well-pled allegations, which must be accepted as true.

A. The Times alleges infringement by end users of Defendants’ products.

Although Microsoft argues that copyright infringement by users of its GenAI products is just a “theoretical possibility” (Mot. at 11), the question at this stage is whether The Times has plausibly alleged that such infringement has taken place. The answer is yes. For example, the Complaint alleges there were “widely publicized” instances “of copyright infringement after ChatGPT, Browse with Bing, and Bing Chat were released.” Compl. ¶ 126. These infringements “actually happened in the real world” (*contra* Mot. at 8)—otherwise, the press could not have reported on them.¹ Similarly, Microsoft downplays the infringing output in the Complaint (Compl.

¹ See, e.g., Kevin Hurler, *OpenAI Pauses ChatGPT’s ‘Browse With Bing’ as Users Bypass Paywalls*, GIZMODO (July 5, 2023), <https://gizmodo.com/openai-pauses-chatgpt-browse-with-bing-paywall-1850605577> (“Less than a week after introducing its integration with Bing, ChatGPT pulled the plug on the collaboration as users discovered they

¶¶ 104, 106, 112, 115, 118, 121, 130, 131, 134) by claiming those infringements resulted from “acrobatics” by “individuals directed by The Times.” Mot. at 11. In reality, the Complaint shows that the prompts given to Defendants’ chatbots simply involved asking for an article’s first paragraph, followed by a request for additional paragraphs. *Id.* ¶¶ 112, 115, 131. Microsoft does not dispute that its products generated such outputs and that they violate The Times’s copyrights.

Microsoft’s reliance on *Matthew Bender & Co. v. West Publishing Co.*, 158 F.3d 693 (2d Cir. 1998) (Mot. at 11), which was decided on summary judgment, highlights the importance of discovery. The copyright holder there “hypothesized” that users of CD-ROM products might try to “replicate” its printed compilations of judicial opinions, a task the court skeptically characterized as “thankless toil.” *Id.* at 706. At summary judgment, the copyright holder could not identify evidence to back up that allegation. *Id.* at 706. If discovery here shows that no chatbot users ever obtained copyrighted content through Defendants’ GenAI products, despite The Times’s allegations showing that this is possible and reports regarding the general public doing so, this claim may not survive summary judgment. That outcome is unlikely, including because the infringing conduct here (obtaining copyrighted content from a chatbot) is nothing like the “thankless toil” that was “hypothesized” in *Bender*. 158 F.3d at 706. But the Court cannot decide this fact-intensive question now, particularly where evidence of user infringements is most likely

could use it to bypass paywalled websites.”); Zainab Choudhry & Aarian Marshall, *News Publishers Are Wary of the Bing Chatbot’s Media Diet*, WIRED (Feb. 11, 2023), <https://www.wired.com/story/news-publishers-are-wary-of-the-microsoft-bing-chatbots-media-diet/> (“Microsoft’s new search interface can serve up key information from articles, removing the need to click—and potentially undermining publisher business models.”); Yona TR Golding, *The News Media and AI: A New Front in Copyright Law*, COLUMBIA JOURNALISM REV. (Oct. 18, 2023), https://www.cjr.org/business_of_news/data-scraping-ai-litigation-lawsuit-artists-authors.php (“In June, ChatGPT Plus users discovered that the chatbot’s internet search feature, Browse with Bing, could even bypass publisher paywalls directly when asked to reprint an article from a URL. The company temporarily disabled the feature ‘in order to do right by content owners,’ which it said it did ‘out of an abundance of caution.’”).

in possession of the Defendants.² See *Arista Recs. LLC v. Usenet.com, Inc.*, 633 F. Supp. 2d 124, 150 (S.D.N.Y. 2009) (granting summary judgment to plaintiff on contributory infringement claim, relying in part on “data from Defendants’ server . . . which constitutes direct evidence of subscribers requesting to download Plaintiffs’ copyrighted works”).

B. The Times alleges Microsoft’s knowledge of infringement.

Microsoft also contends The Times has not adequately alleged Microsoft’s knowledge of infringing outputs, citing out-of-circuit cases to argue that plaintiffs must allege “actual knowledge of specific acts of infringement” or “[w]illful blindness of specific facts.” Mot. at 12 (citations omitted). But that is not the law in this circuit, where “knowledge of specific infringements is not required to support a finding of contributory infringement.”³ *Usenet.com*, 633 F. Supp. 2d at 154; see also *Doe 3*, 604 F.3d at 118 (2d Cir. 2010) (the knowledge requirement is met if defendants “know or have reason to know of the direct infringement”); *Arista Recs., Inc. v. Mp3Board, Inc.*, 2002 WL 1997918, at *7 (S.D.N.Y. Aug. 29, 2002) (Stein, J.) (“A defendant must possess either actual or constructive knowledge of the infringing activity to be found contributorily liable.”).

Even if the higher standard applied, The Times did plead Microsoft’s actual knowledge of infringement. Compl. ¶¶ 126, 180. This allegation is more than plausible given that such

² Nothing in *Tremblay v. OpenAI, Inc.*, 2024 WL 557720, at *2–3 (N.D. Cal. Feb. 12, 2024), leads to a different result. The plaintiffs there alleged that “every output of the OpenAI Language Models is an infringing derivative work,” yet did not explain “what the outputs entail or allege that any particular output is substantially similar—or similar at all—to their books.” *Id.* By contrast, The Times cites numerous examples to show how Defendants’ products output verbatim copies of Times works. Compl. ¶¶ 104, 106, 112, 115, 118, 121, 130, 131, 134.

³ Microsoft’s reference to the Second Circuit’s summary judgment decision in *Gershwin Publishing Corp. v. Columbia Artists Mgmt.*, 443 F.2d 1159 (2d Cir. 1971) (Mot. at 12) does not help it. *Gershwin* did find actual knowledge, but to distort this finding to mean that a plaintiff cannot allege constructive knowledge would mistake the sufficient for the necessary. Moreover, to the extent *Gershwin* is relevant here, it supports The Times. The court found contributory infringement where the defendant, as here, “was in a position to police the infringing conduct [and] derived substantial financial benefit from the actions of the primary infringers.” *Id.* at 1163; see Compl. ¶¶ 55–74. *Viacom International, Inc. v. YouTube, Inc.*, 676 F.3d 19 (2d Cir. 2012) (Mot. at 12) does not help Microsoft either. As Microsoft admits, that case was interpreting the knowledge requirement for the DMCA safe harbor—a statutory interpretation question not implicated by this Motion.

infringements were “widely publicized” (*id.*) and The Times contacted Microsoft to provide notice of observed copyright infringement, *id.* ¶¶ 54, 126; *see ReDigi*, 934 F. Supp. 2d at 658 (reasoning that the defendant “knew or should have known that its service would encourage infringement,” including based on a “cease-and-desist letter” asserting a copyright violation).

The Times’s claim of knowledge is further bolstered by Microsoft’s critical role in developing and overseeing Defendants’ products, which means Microsoft understands how these products use copyrighted works for training and returning synthetic search results in real time. Compl. ¶¶ 73, 108–09; *see Usenet.com*, 633 F. Supp. 2d at 155 (“Defendants knew or should have known of infringement by its users” where they “were explicitly put on notice of the existence of thousands of copies of Plaintiffs’ copyrighted sound recordings available on its service”).

Any dispute Microsoft has with these allegations can be addressed during discovery because a defendant’s assertion that it “lacked knowledge of any infringement . . . is best resolved on summary judgment.” *Snail Games USA Inc. v. Tencent Cloud LLC*, 2022 WL 3575425, at *5 (C.D. Cal. June 6, 2022). “This is reflected by the fact that the cases primarily relied upon by [Microsoft] are summary judgment cases.” *Snail Games*, 2022 WL 3575425, at *5.

Microsoft closes its contributory infringement argument by asserting that, under *Sony*, it cannot be liable “merely based on offering a multi-use product” that has substantial non-infringing uses. Mot. at 13. *Sony* was part of a decades-old line of cases addressing contributory infringement for “distributors of copying equipment” who had a single interaction with their customer at the point-of-sale. *Sony*, 464 U.S. at 421. But the evidence presented in that case showed that Sony did not copy, store, use or distribute copyrighted works to users, nor did it exercise control over users or the content delivered to them once they purchased a VCR. *See id.* at 437–38. The *Sony* Court noted that where, as here, defendants have an “‘ongoing relationship’ with the product or its end-

user,” its reasoning would not apply given the defendants’ ability to control user content. *Usenet.com*, 633 F. Supp. 2d at 156 (quoting *Sony*, 464 U.S. at 438). That is precisely the case here, because Microsoft does much more than just offer a product for one-time purchase: it offers ongoing access to products that are preloaded with copyrighted content, duplicate copyrighted content for real-time synthetic search results, and deliver copyrighted content to users. Compl. ¶¶ 72–73, 95–96, 102, 112–23, 149. Microsoft’s argument that *Sony* bars liability “rides roughshod over a critical part of the Supreme Court’s reasoning in *Sony*.” *Usenet.com*, 633 F. Supp. 2d at 155–56 (reasoning that “Defendants’ service is quite unlike *Sony*” because “Defendants maintain an ongoing relationship with their users”).⁴

II. The Times states DMCA claims on multiple grounds.

The Times alleges that Defendants violated the DMCA by removing The Times’s copyright management information (“CMI”) at two points in time: when building training datasets for the models, Compl. ¶ 184, and during the process of generating outputs from Defendants’ user-facing products like ChatGPT and Copilot, *id.* ¶¶ 185–86. Removal of CMI at both stages implicates the goal of the DMCA to combat intellectual property theft by protecting information that “assist[s] in tracking and monitoring uses of copyrighted works.” S. Rep. No. 105-190 at 16 (1998). The purpose and effect of stripping CMI from The Times’s content during both training and synthetic search retrieval was to stop copyright notices and other CMI from appearing when Defendants distribute The Times’s copyrighted content—thereby concealing Defendants’ infringement. Compl. ¶¶ 125 188, 190.

⁴ See also *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 933 (2005) (rejecting “broad” interpretation of *Sony* that would limit liability “whenever a product is capable of substantial lawful use”); *Napster*, 239 F.3d at 1022 (finding that plaintiffs demonstrated likelihood of success on contributory infringement claim, including because defendant “could block access to the system by suppliers of the infringing material [but] failed to remove the material”); *A&M Records Inc. v. Napster Inc.*, 114 F. Supp. 2d 896, 917 (N.D. Cal. 2000) (“Courts have distinguished the protection *Sony* offers to the manufacture and sale of a device from scenarios in which the defendant continues to exercise control over the device’s use.”).

Microsoft argues that these claims fail because it (i) lacked “culpable awareness that removal of CMI will ‘induce, enable, facilitate, or conceal an infringement’ of The Times’s works,” and (ii) did not remove CMI from “identical copies” of Times works. Mot. at. 14. These arguments ignore key allegations and misconstrue others, all while raising factual disputes that cannot be decided in Microsoft’s favor at this stage.

A. The Times alleges that Microsoft knew or had “reasonable grounds to know” that removal of CMI would conceal or facilitate copyright infringement.

What Microsoft calls the “second scienter requirement” (Mot. at 15) merely requires that Microsoft knew or “had reasonable grounds to know” that removal of CMI would “facilitate” or “conceal” infringement. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 171 (2d Cir. 2020). Fairly read, the Complaint provides ample detail to support its allegation that Microsoft knew or should have known that removal of CMI would “facilitate or conceal” its infringement, both during training and in generating outputs. Compl. ¶ 125 (alleging that Defendants removed CMI during training “with the knowledge that such CMI would not be retained within the models or displayed when the models present unauthorized copies or derivatives of Times Works to users, and thereby facilitate or conceal their infringement”); *see also id.* ¶¶ 124, 188, 190.

To the extent factual questions arise about Microsoft’s knowledge and conduct, they are best resolved after discovery. *See Hirsch v. CBS Broad. Inc.*, 2017 WL 3393845, at *8 (S.D.N.Y. Aug. 4, 2017) (denying motion to dismiss DMCA claim despite “sparse” knowledge allegations because “[c]ourts must be ‘lenient in allowing scienter issues to survive motions to dismiss’” (cleaned up) (quoting *In re DDAVP Antitrust Litig.*, 585 F.3d 677, 693 (2d Cir. 2009))).⁵ This is

⁵ *See also Devocean Jewelry LLC v. Associated Newspapers Ltd.*, 2016 WL 6135662, at *2 (S.D.N.Y. Oct. 19, 2016) (denying motion to dismiss DMCA claim where it was “at least as reasonable to infer that the CMI was altered to conceal or facilitate the alleged infringement as for any other reason”); *Agence France Presse v. Morel*, 769 F. Supp. 2d 295, 305 (S.D.N.Y. 2011) (denying motion to dismiss DMCA claim where arguments about the defendant’s “intent” raised a “fact issue [that] cannot be resolved on a motion to dismiss”).

especially true here because the best evidence of Microsoft’s conduct and knowledge “‘may be found within [Microsoft’s] possession.’” *Argo Contracting Corp. v. Paint City Contractors, Inc.*, 2000 WL 1528215, at *1 (S.D.N.Y. Oct. 16, 2000) (Stein, J.) (citation omitted).

1. Microsoft’s knowledge of the effect of removing CMI during model training

Microsoft argues that, given CMI’s purpose of informing the public that a work is copyrighted, “it makes no sense to base a CMI claim on alleged infringement during training” because “[t]raining takes place out of public view, and training data itself is not disseminated.” Mot. at 15–16. But CMI has a further purpose, as shown by a portion of *Fashion Nova* that Microsoft fails to cite: “The central purpose of CMI is to inform the public that something is copyrighted *and to prevent infringement.*” *Fashion Nova, LLC v. Blush Mark, Inc.*, 2023 WL 4307646, at *4 (C.D. Cal. June 30, 2023) (emphasis added and internal quotation marks omitted).

Permitting liability for the removal of CMI during the LLM training process serves both goals. After being built with copyrighted Times works from which CMI has been removed, the models are built into Defendants’ consumer-facing products and facilitate the output of copyrighted content without CMI. Compl. ¶¶ 78–81; *see also supra* Section I.A (discussing examples of outputs). By stripping CMI during training, Defendants prevent that CMI from being displayed to users as output. *Id.* ¶ 125. The result is that the copyrighted nature of Times content is concealed from users of Defendants’ products. *E.g., id.* ¶¶ 102–07 (providing examples of sample outputs with verbatim excerpts from Times works that omit CMI).

These actions “conceal” and “facilitate” Defendants’ *own* infringement, Compl. ¶¶ 125, 188, 190, rendering Defendants liable under the DMCA. *See Mango*, 970 F.3d at 172 (“[A] defendant’s awareness that distributing copyrighted material without proper attribution of CMI will conceal *his own* infringing conduct satisfies the DMCA’s second scienter requirement.”). And

Microsoft knows this, or at the very least, has “reasonable grounds” to know it. Microsoft has been deeply involved in the training and development of the models, including the selection of datasets used to create them. *See* Compl. ¶ 93 (quoting Microsoft CEO’s statement that Microsoft has “done a lot of work” to “take real care” of training data, including as to its “quality” and “provenance”); *see also id.* ¶ 70 (Microsoft “built” the supercomputer used to train the models). Microsoft knows these models were built from copyrighted content and can directly copy their protected training in their outputs. *Id.* ¶¶ 66–71, 80 & n.10; Mot. at 5.

Because The Times alleges that Defendants removed CMI from copyrighted content during training to conceal and facilitate *their own* infringement, this case is less like *Tremblay*—where plaintiffs advanced only a theory of enabling *third-party* infringement, 2024 WL 557720, at *4—and more like *Shihab v. Complex Media, Inc.*, 2022 WL 3544149, at *5 (S.D.N.Y. Aug. 17, 2022), in which the court refused to dismiss a DMCA claim alleging the defendant “knew or should have known” that its removal of CMI “would conceal its contemporaneous [] infringement.”

Microsoft concludes with the absurd argument that The Times “concedes away any claim” because Paragraph 8 of the Complaint “acknowledg[es] Microsoft’s strong belief that ‘thi[s] conduct is protected as fair use.’” Mot. at 16. It does not. What that paragraph alleges—but Microsoft does not quote—is that “[p]ublicly, Defendants insist that their conduct is protected as ‘fair use,’” and the Complaint goes on to explain why that position is incorrect. Compl. ¶ 8 (emphasis added). In any event, even if there were a mistake-of-law defense to liability (which there is not) discovery would be needed to determine whether Microsoft believes what it professes.

2. Microsoft’s knowledge of the effect of removing CMI from generative outputs

In addition to the training-based violations described above, Defendants also violated section 1202(b)(1) of the DMCA by “remov[ing] The Times’s [CMI]” in the process of

“generating outputs” from Defendants’ user-facing products like ChatGPT and Copilot. Compl. ¶¶ 185–86; 17 U.S.C. § 1202(b)(1). These outputs also trigger liability under section 1202(b)(3) because Defendants are knowingly distributing copies of Times works without CMI. *See* 17 U.S.C. § 1202(b)(3). This claim is not “nonsensical” (Mot. at 17) but rather an example of the type of conduct at the core of what these DMCA provisions prohibit: the removal of CMI from a copyrighted work that is delivered to users. *See Mango*, 970 F.3d at 169 (describing claim under § 1202(b)(3) where “defendant distributed copyrighted work” from which CMI was removed).

Like its attack on the contributory infringement claim, Microsoft relies primarily on its own *ipse dixit* that people in the “real-world” do not use their products to infringe copyrights. Mot. at 16. That assertion cannot be credited on a motion to dismiss, and The Times amply alleges otherwise. Compl. ¶ 126 (“Defendants were aware of many examples of copyright infringement after ChatGPT, Browse with Bing, and Bing Chat were released, some of which were widely publicized.”); *see also supra* n.1 (citing multiple examples of public reports explaining how Defendants’ products are used to serve up paywall-protected content). The Complaint also includes numerous examples of Defendants’ products generating infringing outputs in response to user queries. Compl. ¶¶ 104–22. These allegations are more than sufficient to plead actual “aware[ness],” Compl. ¶ 126; or at a minimum, that Microsoft “should have known” that its products were generating infringing outputs, *see Shihab*, 2022 WL 3544149, at *5 (DMCA claim can be based on constructive knowledge).

Microsoft’s reliance on *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 673 (9th Cir. 2018), is misplaced. Mot. at 17. As Microsoft acknowledges, the plaintiffs in *Stevens* were focused on whether removal of CMI would facilitate future infringement by “a third-party.” Mot. at 17. Here, however, the Complaint alleges that Defendants’ removal of CMI, both during training and also

in generating outputs, conceals and facilitates *their own* infringement. Compl. ¶¶ 125, 188, 190; *see Mango*, 970 F.3d at 174 (distinguishing *Stevens* because it “did not address whether a defendant’s own infringement satisfies Section 1202(b)(3)’s second scienter requirement”). *Mango* then confirmed the viability of The Times’s theory. *See Mango*, 970 F.3d at 169 (holding “that the DMCA does not require [] evidence” of “future, third-party infringement”).

Finally, Microsoft’s assertion that its removal of CMI is “irrelevant” if the user still “knows where the article is from” (Mot. at 17) makes no sense. Simply knowing that something was published by The Times does not mean a user knows it is subject to copyright protection, which is why The Times publishes that CMI on every page of its website—after all, if you’re on The Times’s website, you also know the origin of the content. Moreover, the examples in the Complaint are illustrative, not exhaustive, and presumably the “widely publicized” examples of copyright infringement involved other methods to obtain copyrighted content that do not require the user to know the source. Compl. ¶ 126. Indeed, the Complaint offers one such example in which a prompt returned Times content without specifically referencing The Times. *Id.* ¶ 106. Microsoft’s contrary arguments raise a dispute regarding the factbound question of knowledge, which cannot be resolved on the pleadings. *See supra* Section II.A (citing cases).

B. The Times alleges removal of CMI from copies of Times works.

Microsoft’s main argument regarding removal of CMI is that Defendants have not copied enough of The Times’s works to be liable for stripping their CMI. Mot. at 17–19. This argument relies on a fundamental misreading of the Complaint. At all three steps—compiling training datasets, compiling synthetic search results, and preparing end-user outputs—The Times has alleged that Defendants copied entire articles and removed CMI from them. Compl. ¶¶ 184–86.

That alone distinguishes the cases relied upon by Microsoft, many decided at summary judgment rather than a motion to dismiss, finding no liability for failure to copy or transpose CMI where the offenders copied only “aspects” or “elements” of a copyrighted work. Where the defendant copied the entire work and removed CMI from it during the process, the DMCA does not require it to subsequently display the entire work in order to be liable. *See, e.g., Planck LLC v. Particle Media, Inc.*, 2021 WL 5113045, at *6 (S.D.N.Y. Nov. 3, 2021) (denying motion to dismiss the DMCA claim where “Defendants [] distributed infringing excerpts from Plaintiff’s stories” from which CMI had been removed).

For synthetic search results and outputs containing verbatim excerpts of Times works, which are actionable under both section 1202(b)(1) and (3), Microsoft argues that, so long as the output is not identical to the original work, there is no claim. Mot. at 19. This is yet another inappropriate attempt to argue the facts. Microsoft did not just copy “a page” (Mot. at 20); the Complaint alleges that Defendants copied the equivalent of the whole book—here, complete articles—including the copyright page, and are now distributing substantial portions of that “book” with the copyright page omitted. Compl. ¶ 189 (“Defendants created copies and derivative works based on The Times’s works [and] distribut[ed] these works without their copyright-management information.”); *see also id.* ¶ 185 (Defendants “removed The Times’s [CMI] through generating synthetic search results, including removing The Times’s [CMI] when scraping Times Works from The Times’s websites and generating copies or derivatives of Times Works as output for the Browse with Bing and Bing Chat offerings.”); *id.* ¶ 102 (“synthetic search results [] are substantially similar to Times Works generated from copies stored in Bing’s search index”); *id.* ¶ 108 (describing the process by which Defendants’ LLMs make copies and “stitch together paraphrases or quotes from the copied Times content”). These allegations must be accepted as true

at this stage. That all but one of Microsoft’s cases were decided on summary judgment, and not a motion to dismiss, underscores why the Motion should be denied.⁶

The cases cited by Microsoft concerning “fail[ure] to transpose CMI” from one work into another” are similarly inapposite, as they involved the copying of a single design element or minor portion of the work.⁷ Mot. at 19. The Times does not seek to hold Defendants accountable for neglecting to “transpose” CMI into a portion of an article that otherwise lacked CMI; it seeks to hold Defendants accountable for removing CMI from the copied articles and then distributing those articles, including substantial portions or derivative works created from them, to users, knowing that CMI had been removed.

Finally, Microsoft does not meaningfully challenge The Times’s allegations regarding removal of CMI for the training-based portion of the claim (Mot. at 20), and it is clear why. The Complaint contains detailed allegations describing removal of CMI from copies of Times works before and during the training process. Compl. ¶¶ 84–89. The Complaint also alleges that Defendants removed CMI from “Times Works scraped directly from The Times’s websites.” *Id.* ¶ 184. There is no need for these datasets or training materials to be “disseminated” (Mot. at 16),

⁶ See *Design Basics, LLC v. WK Olson Architects, Inc.*, 2019 WL 527535, at *1 (N.D. Ill. Feb. 11, 2019) (no violation where defendant “only . . . copied elements of” architectural design plans); *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609 (S.D.N.Y. 2018) (summary judgment decision finding no violation because the allegedly infringing advertisement “b[ore] no resemblance” to the copied work); *Falkner v. Gen. Motors LLC*, 393 F. Supp. 3d 927, 938 (C.D. Cal. 2018) (summary judgment decision finding no violation where it was “undisputed” that no one “removed or altered the copyright management information that Bernstein did include in his photograph,” and where the photographs captured portions of the works that lacked any CMI); *Faulkner Press, L.L.C. v. Class Notes, L.L.C.*, 756 F. Supp. 2d 1352, 1359 (N.D. Fla. 2010) (summary judgment decision finding no violation where students took notes based on a copyrighted textbook); *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) (summary judgment decision finding no violation where CMI did “not appear in the images” that defendants copied).

⁷ See *Dolls Kill, Inc. v. Zoetop Bus. Co.*, 2022 WL 16961477, at *3 (C.D. Cal. Aug. 25, 2022) (no violation for omission of CMI in “knockoff products”); *A’Lor Int’l, Ltd. v. Tappers Fine Jewelry*, 2012 WL 12921035, at *1 (C.D. Cal. Aug. 8, 2012) (no violation for omitting CMI from “copies of [] jewelry”); *Frost-Tsuji Architects v. Highway Inn, Inc.*, 2015 WL 263556, at *2 (D. Haw. Jan. 21, 2015) (no violation where defendant “developed its own drawing based on [plaintiff’s] work”). *Frost-Tsuji* is also inapposite because it is yet another summary judgment decision.

because The Times’s training-based claim is brought under § 1202(b)(1)’s prohibition on removal of CMI, not the “distribution” of works where CMI has been removed.

C. There is no “judicial consensus” that DMCA claims cannot be brought against GenAI developers.

Microsoft falsely asserts there is “an emerging judicial consensus” that plaintiffs cannot bring DMCA claims for removing CMI from LLM training data and outputs. Mot. at 14. But the three cases Microsoft cites for this proposition all involved pleading failures that the plaintiffs were permitted to amend, none of which are present here. *See Andersen v. Stability AI Ltd.*, 2023 WL 7132064, at *11 (N.D. Cal. Oct. 30, 2023) (finding there was “nothing in the Complaint” to describe the CMI that was removed, making the allegation “wholly conclusory”); *Tremblay*, 2024 WL 557720, at *5 (plaintiffs alleged “no facts” to support their assertion that CMI was removed during training); *Doe 1 v. GitHub, Inc.*, 2024 WL 235217, at *9 (N.D. Cal. Jan. 22, 2024) (dismissing an amended DMCA claim which alleged that Copilot’s output “is often a modification” of their works, “as opposed to an ‘identical copy’”).

More instructive is the first decision in *Doe 1 v. GitHub, Inc.*, 672 F. Supp. 3d 837, 859 (N.D. Cal. 2023), in which the court allowed plaintiffs’ DMCA claim to proceed based on similar allegations of verbatim copying that The Times makes here. As the *Tremblay* court acknowledged, the *GitHub* plaintiffs stated a DMCA claim because they alleged “where the CMI typically appeared, that defendants were aware that the CMI appeared repeatedly, and that defendants subsequently ‘trained these programs to ignore or remove CMI and therefore stop producing it,’” which was “sufficient to support a ‘reasonable inference that Defendants intentionally designed the programs to remove CMI.’” *Tremblay*, 2024 WL 557720, at *4 (quoting *GitHub Inc.*, 672 F. Supp. 3d at 858). So too here. The Times has identified its specific CMI (Compl. ¶¶ 125, 182, 187), alleged Microsoft’s knowledge of the removal of that CMI (*id.* ¶¶ 72, 125, 184–87), and

included more than 100 examples of infringing, verbatim output displayed in Defendants’ products without such CMI (*id.* Ex. J; ¶¶ 4–5, 98–122, 130.). To the extent three opinions from the same district qualify as a “consensus” on when DMCA claims predicated on LLMs should survive a motion to dismiss, The Times’s Complaint easily passes the test.

III. The Copyright Act does not preempt The Times’s “hot-news” unfair competition by misappropriation claim.

The Times asserts a “hot-news” unfair competition by misappropriation claim as recognized in *International News Service v. Associated Press*, 248 U.S. 215 (1918) (“*INS*”). As Congress has explained, 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.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). Like OpenAI, Microsoft acknowledges that the Copyright Act does not preempt such a claim but argues that it does not apply to these facts. Mot. at 21. Like OpenAI, Microsoft is wrong.

An *INS*-type claim requires “free-riding” by a defendant where the “defendant was taking news gathered and in the process of dissemination by the [plaintiff] and selling that news as though the defendant itself had gathered it.” *Barclays*, 650 F.3d at 902–03. There are two different formulations of a multi-part test articulated in *NBA* to assert an *INS*-type non-preempted claim, one of which Microsoft has cited in its Motion: “(i) the time-sensitive value of factual information;

(ii) the free-riding by a defendant; and (iii) the threat to the very existence of the product or service provided by the plaintiff.”⁸ *Id.* at 900; Mot. at 22.

Here, The Times has pled a non-preempted *INS*-type claim. The Times collects and disseminates time-sensitive content at great cost, including news and product or service recommendations. Compl. ¶¶ 32–37, 193. Defendants are free-riding by stealing The Times’s time-sensitive content generated at significant expense, and then, in direct competition with The Times, distributing that content to people who use Defendants’ products like ChatGPT and Copilot. The Complaint explains how these products “use Microsoft’s Bing search index . . . to generate responses that contain verbatim experts and detailed summaries of Times articles that . . . undermine and damage The Times’s relationship with its readers and deprive The Times of subscription, licensing, advertising, and affiliate revenue.” *Id.* ¶ 5. The “synthetic” search results generated by Defendants’ products purport to answer user queries directly, resulting in Defendants keeping users within their own sites and applications instead of referring users to The Times, as organic search results would do. *Id.* ¶ 72; *see also id.* ¶¶ 108–23 (“Unauthorized Dissemination of Current News”); *id.* ¶¶ 193–197 (Defendants are “free-riding on The Times’s significant efforts and investment of human capital to gather this information”). Defendants’ activities substantially threaten The Times by, *inter alia*, “divert[ing] important traffic away,” *id.* ¶ 110, and “creat[ing] less incentive for users to navigate” to The Times’s websites, *id.* ¶ 129. *See id.* ¶¶ 52, 108–23.

⁸ As explained in *Barclays*, the multi-part tests in *NBA* “were sophisticated observations in aid of the Court’s analysis of the difficult preemption issues presented to it,” but “could not all be equivalent to a statutory command to which we or the district court are expected to adhere.” 650 F.3d at 901. Thus, while it is unclear whether a court must use this particular test, The Times does not dispute its application because its “hot news” claim would survive regardless of which test applies.

Microsoft presents two separate challenges to The Times’s *INS*-type claim, first with respect to news content and second for its recommendations content, including that published by Wirecutter. Both challenges fail.

A. The Times states a non-preempted *INS*-type claim with respect to news content.

Microsoft raises factual disputes about whether the Defendants’ Bing-powered tools disseminate “time-sensitive news precisely at the point where the profit is to be reaped” and sufficiently threaten The Times’s business. Mot. at 22–23. The Times sufficiently alleges both.

The Times alleges that it engages in breaking ⁹ news reporting and that the Defendants misappropriate The Times’s current news as part of the Bing-powered products’ generation of synthetic search results. *See, e.g.*, Compl. ¶¶ 32–37, 193. The Complaint also provides multiple specific examples of The Times’s breaking news content appearing in Defendants’ synthetic search results, such as Hamas’s attack on Israel (*id.* ¶ 112) and a woman being shoved in front of a subway train (*id.* ¶¶ 118–20). *See generally id.* ¶¶ 108–23. Accordingly, The Times has satisfied its burden at this stage to “show that [it has] a plausible claim for hot news misappropriation” by providing one or more examples of “hot information” having been misappropriated. *BanxCorp*, 723 F. Supp. 2d at 613 (denying motion to dismiss hot news misappropriation claim).

Microsoft argues that the Complaint contains only “naked assertions that GPT-based tools directly compete with Times content.” Mot. at 22. This is simply not correct. The Complaint contains specific allegations regarding precisely how the Defendants’ GPT-based synthetic search products steal The Times’s current news, spit out substantially similar copies of such news, and do so in a manner that makes users less likely to navigate to The Times’s own website. Compl. ¶¶ 108–10. The Complaint also alleges that the Defendants’ synthetic search results based on The

⁹ Microsoft argues that the Complaint does not define “time-sensitive” news, Mot. at 22, but fails to explain how The Times’s news—which is reported in print on a daily basis and online in near real-time—does not meet the definition.

Times’s current news content threaten The Times’s business precisely at the point where the profit is to be reaped by, *inter alia*, “keep[ing] users within [Defendants’] own search ecosystem” (*id.* ¶ 52), “divert[ing] important traffic away” from The Times (*id.* ¶ 110), “creat[ing] less incentive for users to navigate” to The Times’s websites (*id.* ¶ 129), and “depriving The Times of the opportunity to receive referral revenue” (*id.* ¶ 194). *See also* Compl. ¶¶ 4–5, 52, 72, 108–23, 129. These allegations are sufficient to meet the pleading requirements for a hot news misappropriation claim. *See, e.g., BanxCorp*, 723 F. Supp. 2d at 613–14 (finding direct competition element satisfied “because it is supported by the factual allegation that Plaintiffs market HYSAs and CDs that compete with Defendants’ co-branded savings accounts”).

Microsoft also disputes The Times’s allegations concerning “the quantity of copying” and “the immediacy of distribution necessary to sustain a ‘hot’ news claim” (Mot. at 22), but such factual disputes cannot be resolved at this stage. *DiBlasio v. Novello*, 344 F.3d 292, 304 (2d Cir. 2003). This point is made clear by the very case—and the specific passage from that case—upon which Microsoft relies in support of its argument. Specifically, Microsoft’s motion states: “The Complaint therefore alleges ‘neither the quantity of copying nor the immediacy of distribution necessary to sustain a ‘hot’ news claim.’ *Fin. Info., Inc. v. Moody’s Invs. Serv., Inc.*, 808 F.2d 204, 209 (2d Cir. 1986).” Mot. at 22. However, Microsoft has elided the key language from the passage in *Moody’s* reflecting that the ruling was not made on a motion to dismiss, but instead only after a full evidentiary record had been made at trial: “FII proved neither the quantity of copying nor the immediacy of distribution necessary to sustain a ‘hot’ news claim.” *Moody’s*, 808 F.2d at 209 (emphasis added). *Barclays*, too, was decided only after a full record had been developed at trial.¹⁰

¹⁰ Microsoft’s string citation to cases in which courts found hot news claims preempted all involve facts entirely different than those alleged in the Complaint. *See Greer v. Fox News Media*, 2023 WL 2671796, at *1 (2d Cir. Mar. 29, 2023) (affirming a finding that the “time-sensitive” prong was not met where the plaintiff recognized a “time lag”

B. The Times states a non-preempted *INS*-type claim with respect to its recommendations.

As to The Times’s product and services recommendations, Microsoft argues that: (1) the Complaint “ventures no explanation of how its product reviews are ‘time sensitive,’” Mot. at 23, and (2) the “free-riding” alleged in the complaint is preempted under *Barclays* because Defendants are not “sell[ing] Wirecutter reviews as their own,” Mot. at 24. Both arguments miss the mark.

First, the Complaint alleges that “Wirecutter likewise compiles and produces time-sensitive recommendations for readers.” Compl. ¶ 193. Microsoft does not dispute the plausibility of this allegation—nor could it. Wirecutter often provides time-sensitive reviews in response to critical world events,¹¹ time-sensitive sales,¹² or “daily deals.”¹³

Second, the facts alleged in the Complaint are analogous to the hypothetical *INS*-type claim the Second Circuit identified in *Barclays* as *not* preempted: “[i]f a Firm were to collect and disseminate . . . facts about securities recommendations in the brokerage industry . . . and [defendant] were to copy the facts contained in the Firm’s hypothetical service, it might be liable to the Firm on a ‘hot-news’ misappropriation theory.” 650 F.3d at 905–06; *see also NBA*, 105 F.3d at 854 (“[I]f appellants in the future were to collect facts from an enhanced Gamestats pager to retransmit them to SportsTrax pagers, that would constitute free-riding and might well cause Gamestats to be unprofitable because it had to bear costs to collect facts that SportsTrax did not.”).

between his reporting and the purported misappropriation and argued that “there is no expiration date for hot news”); *ML Genius Holdings LLC v. Google LLC*, 2022 WL 710744, at *5 (2d Cir. Mar. 10, 2022) (finding that the “time-sensitive” prong was not adequately alleged where the plaintiff argued that its business model depended on timeliness but did not argue that the misappropriated information was time-sensitive); *DBW Partners, LLC v. Mkt. Sec., LLC*, 2023 WL 2610498, at *5 (D.D.C. Mar. 23, 2023) (holding that “free-riding” was not alleged where plaintiff argued that its “analysis” of the market news was misappropriated by the defendant).

¹¹ *See, e.g.*, Tim Heffernan & Ellen Airhart, *How Can I Clear My Home of Wildfire Smoke*, WIRECUTTER (June 7, 2023), <https://www.nytimes.com/wirecutter/blog/air-purifier-wildfire-smoke/>.

¹² *See, e.g.*, Rose Maura Lorre, *The Best Amazon Prime Day 2021 Deals Still Available on Day 2*, WIRECUTTER (June 22, 2021), <https://www.nytimes.com/wirecutter/money/best-prime-day-deals-2021-day-2/>.

¹³ *See Daily Deals*, WIRECUTTER, <https://www.nytimes.com/wirecutter/deals/>.

Moreover, the “free riding” alleged here is entirely distinguishable from that at issue in *Barclays*. In *Barclays*, the Second Circuit grounded its ruling on its conclusion that the defendant, Fly, itself was “collecting, collating, and disseminating” the newsworthy event that a brokerage firm had recommended a particular stock—a fact that was likely to cause the price of the recommended stock to go up. *Barclays*, 650 F.3d at 903 (“It is Fly’s accurate attribution of the Recommendation to the creator that gives this news its value.”); *id.* at 905. In other words, the mere fact that a brokerage firm recommended a particular stock was itself a newsworthy event that could be reported on (by other humans).

Microsoft does not argue that the fact that The Times publishes particular content is a newsworthy event on which its models somehow report as part of its ingestion and output of all (or virtually all) Times content. Microsoft also ignores entirely the allegation that it strips affiliate links from Wirecutter product reviews, Compl. ¶¶ 128, 132, a fact that was not present in *Barclays* and one that underscores that this claim is distinct from The Times’s copyright claims. Contrary to Microsoft’s argument that “the Complaint offers no real-world facts suggesting meaningful diversion of revenue from Wirecutter to Defendants,” Defendants stand to reap a portion of The Times’s revenues, gained from affiliate links or otherwise, by “keep[ing] users within their own search ecosystem.”¹⁴ See Compl. ¶ 52.

CONCLUSION

This Court should deny the motion in its entirety. Alternatively, the Court should grant leave to amend for any claim, particularly because The Times has not yet amended.

¹⁴ Microsoft includes in its motion (at p. 25) a string cite of inapposite cases, most of which do not even analyze the elements of the tests set out in *NBA*. See *Gogo Apparel, Inc. v. True Destiny, LLC*, 2020 WL 5578336, at *3 (S.D.N.Y. Sept. 17, 2020) (no analysis of *NBA* tests); *Adina’s Jewels, Inc. v. Shashi, Inc.*, 442 F. Supp. 3d 766, 772 (S.D.N.Y. 2020) (same); *Orange Cnty. Choppers, Inc. v. Olaes Enterprises, Inc.*, 497 F. Supp. 2d 541, 556 (S.D.N.Y. 2007) (same); see also *DBW Partners*, 2023 WL 2610498, at *5 (holding that “free-riding” was not alleged where plaintiff argued that its “analysis” of the market news had been misappropriated by the defendant).

Dated: March 18, 2024

/s/ Ian Crosby

Ian Crosby (*admitted pro hac vice*)
Genevieve Vose Wallace (*admitted pro hac vice*)
Katherine M. Peaslee (*pro hac vice pending*)
SUSMAN GODFREY L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 516-3880
Facsimile: (206) 516-3883
icrosby@susmangodfrey.com
gwallace@susmangodfrey.com
kpeaslee@susmangodfrey.com

Davida Brook (*admitted pro hac vice*)
Emily K. Cronin (*admitted pro hac vice*)
Ellie Dupler (*admitted pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Ave of the Stars, Suite 1400
Los Angeles, CA 90067
Telephone: (310) 789-3100
Facsimile: (310) 789-3150
dbrook@susmangodfrey.com
ecronin@susmangodfrey.com
edupler@susmangodfrey.com

Elisha Barron (5036850)
Zachary B. Savage (ZS2668)
Tamar Lusztig (5125174)
Alexander Frawley (5564539)
Eudokia Spanos (5021381)
SUSMAN GODFREY L.L.P.
1301 Avenue of the Americas, 32nd Floor
New York, NY 10019
Telephone: (212) 336-8330
Facsimile: (212) 336-8340
ebarron@susmangodfrey.com
zsavage@susmangodfrey.com
tlusztig@susmangodfrey.com
afrawley@susmangodfrey.com
espanos@susmangodfrey.com

Scarlett Collings (*admission pending*)
SUSMAN GODFREY L.L.P.
1000 Louisiana, Suite 5100
Houston, TX 77002
Telephone: (713) 651-9366
Facsimile (713) 654-6666

scollings@susmangodfrey.com

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

Attorneys for Plaintiff
The New York Times Company