

**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 OPENAI
DEFENDANTS' PARTIAL MOTION TO DISMISS (DKT. 51)**

TABLE OF CONTENTS

- I. INTRODUCTION 1
- II. BACKGROUND 2
 - A. The New York Times and its business model are built on world-class journalism. 2
 - B. OpenAI and its business model are built on mass copyright infringement. 3
 - C. Defendants’ user-facing products further infringe The Times’s copyrights..... 4
 - D. The Times files suit to protect its works. 5
- III. LEGAL STANDARD 6
- IV. ARGUMENT 6
 - A. The Times’s direct infringement claim is timely..... 6
 - B. The Times states a contributory infringement claim..... 9
 - C. The Times states DMCA claims on multiple grounds. 14
 - 1. The Times identifies the CMI at issue and alleges it is conveyed with its works..... 15
 - 2. The Times states a DMCA claim under § 1202(b)(1)..... 16
 - a) The Times states a § 1202(b)(1) claim based on training 16
 - b) The Times states a § 1202(b)(1) claim based on outputs..... 19
 - 3. The Times states a DMCA claim under § 1202(b)(3)..... 20
 - 4. The Times has standing to sue under the DMCA. 20
 - D. The Times’s “hot-news” unfair competition by misappropriation claim is not preempted by the Copyright Act..... 21
- V. CONCLUSION..... 25

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>A&M Recs., Inc. v. Napster, Inc.</i> , 239 F.3d 1004 (9th Cir. 2001)	12
<i>Agence France Presse v. Morel</i> , 769 F. Supp. 2d 295 (S.D.N.Y. 2011).....	16, 18
<i>Andersen v. Stability AI Ltd.</i> , 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023).....	15
<i>Argo Contracting Corp. v. Paint City Contractors, Inc.</i> , 2000 WL 1528215 (S.D.N.Y. Oct. 16, 2000).....	11
<i>Arista Recs. LLC v. Usenet.com, Inc.</i> , 633 F. Supp. 2d 124 (S.D.N.Y. 2009).....	10, 12
<i>Arista Recs., Inc. v. Flea World, Inc.</i> , 2006 WL 842883 (D.N.J. Mar. 31, 2006).....	10
<i>Arista Recs., Inc. v. Mp3Board, Inc.</i> , 2002 WL 1997918 (S.D.N.Y. Aug. 29, 2002).....	13
<i>Arista Recs., LLC v. Doe 3</i> , 604 F.3d 110 (2d Cir. 2010).....	9, 13
<i>BanxCorp v. Costco Wholesale Corp.</i> , 723 F. Supp. 2d 596 (S.D.N.Y. 2010).....	6, 16, 17
<i>Barclays Cap. Inc. v. Theflyonthewall.com, Inc.</i> , 650 F.3d 876 (2d Cir. 2011).....	22, 23, 24, 25
<i>BMG Rights Mgmt. (US) LLC v. Cox Commc'ns, Inc.</i> , 881 F.3d 293 (4th Cir. 2018)	14
<i>Calcutti v. SBU, Inc.</i> , 273 F. Supp. 2d 488 (S.D.N.Y. 2003).....	8
<i>Capitol Records, LLC v. ReDigi Inc.</i> , 934 F. Supp. 2d 640 (S.D.N.Y. 2013).....	9, 10, 12
<i>Cortec Indus., Inc. v. Sum Holding L.P.</i> , 949 F.2d 42 (2d Cir. 1991).....	25

Devocean Jewelry LLC v. Associated Newspapers Ltd.,
2016 WL 6135662 (S.D.N.Y. Oct. 19, 2016).....17, 18

DiBlasio v. Novello,
344 F.3d 292 (2d Cir. 2003).....24

Doe 1 v. GitHub, Inc.,
2023 WL 3449131 (N.D. Cal. May 11, 2023).....17

Faulkner v. Beer,
463 F.3d 130 (2d Cir. 2006).....8

Financial Info., Inc. v. Moody’s Invs. Serv., Inc.,
808 F.2d 204 (2d Cir. 1986).....24

Fischer v. Forrest,
286 F. Supp. 3d 590 (S.D.N.Y. 2018).....19

Granite Partners, L.P. v. Bear, Stearns & Co.,
58 F. Supp. 2d 228 (S.D.N.Y. 1999).....11

Grimes v. N.Y. & Presbyterian Hosp.,
2024 WL 816208 (S.D.N.Y. Feb. 26, 2024).....6

Hartmann v. Apple, Inc.,
2021 WL 4267820 (S.D.N.Y. Sept. 20, 2021).....14

Hartmann v. Popcornflix.com LLC,
2023 WL 5715222 (S.D.N.Y. Sept. 5, 2023).....12

Hesse v. Godiva Chocolatier, Inc.,
463 F. Supp. 3d 453 (S.D.N.Y. 2020).....7

Hirsch v. CBS Broad. Inc.,
2017 WL 3393845 (S.D.N.Y. Aug. 4, 2017).....18

Hirsch v. Rehs Galleries, Inc.,
2020 WL 917213 (S.D.N.Y. Feb. 26, 2020).....7

In re DDAVP Antitrust Litig.,
585 F.3d 677 (2d Cir. 2009).....18

Int’l News Serv. v. Associated Press,
248 U.S. 215 (1918).....21

Janik v. SMG Media, Inc.,
2018 WL 345111 (S.D.N.Y. Jan. 10, 2018).....15

Kelly v. Arriba Soft Corp.,
77 F. Supp. 2d 1116 (C.D. Cal. 1999)21

Luvdarts, LLC v. AT & T Mobility, LLC,
710 F.3d 1068 (9th Cir. 2013)14

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

McGlynn v. Sinovision Inc.,
2024 WL 643021 (S.D.N.Y. Feb. 15, 2024).....7

Miller v. Netventure24 LLC,
2021 WL 3934262 (S.D.N.Y. Aug. 6, 2021).....21

Nat’l Basketball Ass’n v. Motorola, Inc.,
105 F.3d 841 (2d Cir. 1997).....22, 23

Olusola v. Don Coqui Holding Co., LLC,
2021 WL 631031 (E.D.N.Y. Feb. 18, 2021).....21

Parisienne v. Scripps Media, Inc.,
2021 WL 3668084 (S.D.N.Y. Aug. 17, 2021).....8

Perfect 10, Inc. v. Amazon.com, Inc.,
508 F.3d 1146 (9th Cir. 2007)20

Pierson v. Infinity Music & Ent., Inc.,
300 F. Supp. 3d 390 (D. Conn. 2018).....15

Pilla v. Gilat,
2020 WL 1309086 (S.D.N.Y. Mar. 19, 2020)19

PK Music Performance, Inc. v. Timberlake,
2018 WL 4759737 (S.D.N.Y. Sept. 30, 2018).....7, 9

Planck LLC v. Particle Media, Inc.,
2021 WL 5113045 (S.D.N.Y. Nov. 3, 2021).....19

Reilly v. Plot Commerce,
2016 WL 6837895 (S.D.N.Y. Oct. 31, 2016).....20, 21

Roberts v. BroadwayHD LLC,
518 F. Supp. 3d 719 (S.D.N.Y. 2021).....19

Rosen v. Amazon.com, Inc.,
2014 WL 12597073 (C.D. Cal. May 28, 2014)11

Shihab v. Complex Media, Inc.,
2022 WL 3544149 (S.D.N.Y. Aug. 17, 2022).....18

Snail Games USA Inc. v. Tencent Cloud LLC,
2022 WL 3575425 (C.D. Cal. June 6, 2022)11

Sohm v. Scholastic Inc.,
959 F.3d 39 (2d Cir. 2020).....6

State St. Glob. Advisors Tr. Co. v. Visbal,
431 F. Supp. 3d 322 (S.D.N.Y. 2020).....14

Steele v. Bongiovi,
784 F. Supp. 2d 94 (D. Mass. 2011)21

Tiffany (NJ) Inc. v. eBay Inc.,
600 F.3d 93 (2d Cir. 2010).....10

Tremblay v. OpenAI, Inc.,
2024 WL 557720 (N.D. Cal. Feb. 12, 2024)17, 19, 20

Viacom Int’l, Inc. v. YouTube, Inc.,
676 F.3d 19 (2d Cir. 2012).....14

Wright v. Miah,
2023 WL 6219435 (E.D.N.Y. Sept. 7, 2023)19

Zuma Press, Inc. v. Getty Images (US), Inc.,
845 F. App’x 54 (2d Cir. 2021)18

Statutes

17 U.S.C. § 507(b)6

17 U.S.C. § 120214, 15, 16, 17, 19, 20

Other Authorities

Bing Chatbot’s Media Diet, WIRED (Feb. 11, 2023)10

Brown et al., *Language Models are Few-Shot Learners* (2020).....13

H.R. No. 94-1476, reprinted in 1976 U.S.C.C.A.N.21

John Branch, *Snow Fall: The Avalanche at Tunnel Creek*, THE NEW YORK TIMES
(Dec. 20, 2012)15

Kevin Hurler, *OpenAI Pauses ChatGPT’s ‘Browse With Bing’ as Users Bypass
Paywalls*, GIZMODO (July 5, 2023)10

Users Bypass Paywalls, GIZMODO (July 5, 2023)10

Yona TR Golding, *The News Media and AI: A New Front in Copyright Law*,
COLUMBIA JOURNALISM REVIEW (Oct. 18, 2023).....11

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

I. INTRODUCTION

Lacking any real grounds for dismissal, OpenAI devotes much of its filing to grandstanding about issues on which it hasn't moved. Its Motion introduces no fewer than 19 extrinsic documents, none of which can be properly considered on a motion to dismiss, in a submission that for nearly 10 pages reads more like spin than a legal brief.

Conspicuously, OpenAI's attention-grabbing claim that The Times "hacked" its products (Mot. at 2) is as irrelevant as it is false. As Exhibit J to the Complaint makes clear, The Times elicited examples of memorization by prompting GPT-4 with the first few words or sentences of Times articles. That work was only necessary because OpenAI does not disclose the content it uses to train its models and power its user-facing products. Yet in OpenAI's telling, The Times engaged in wrongdoing by detecting OpenAI's theft of The Times's own copyrighted content. OpenAI's true grievance is not about how The Times conducted its investigation, but instead what that investigation exposed: that Defendants built their products by copying The Times's content on an unprecedented scale—a fact that OpenAI does not, and cannot, dispute.

Despite seeking to justify this conduct however it can, OpenAI does not move to dismiss the lead claim that it infringed The Times's copyrights to train and operate its latest models. Against those claims it does challenge, OpenAI advances mainly factual arguments that cannot be decided on the pleadings. *First*, to support its statute-of-limitations argument for claims based on models developed before December 2020, OpenAI asks this Court to make a factual finding that the makeup of the datasets used to train those models was "common knowledge" in 2020—even though OpenAI's "viral" ChatGPT chatbot was not released until November 2022. Mot. at 9. *Second*, OpenAI's bid to dismiss the contributory infringement claim turns on disputed facts about user behavior and would require the Court to accept its assertion that "[i]n the real world, people do not use ChatGPT or any other OpenAI product for that purpose" (Mot. at 1)—despite the widely

reported use of ChatGPT to bypass paywalls. *Third*, OpenAI’s attack on the DMCA claim turns on disputed fact issues concerning the “design[]” of OpenAI’s model-training process, Mot. at 19, which cannot be resolved before discovery into that design. *Fourth*, OpenAI asks the Court to dismiss the “hot news” unfair competition by misappropriation claim by ignoring The Times’s allegations of Defendants’ free-riding and deciding that The Times’s product recommendations are not generated by “efforts akin to reporting,” Mot. at 24, yet another premature argument. Discovery, not dismissal, is warranted to resolve each of these well-pleaded claims.

II. BACKGROUND

A. 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, metadata, 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. Generating and maintaining traffic to The Times’s content is critical to its revenue streams. *Id.* ¶ 41. To facilitate that traffic, The Times permits search engines like Microsoft Bing to access and index its content, but inherent in this value exchange is the idea that search engines will direct users to The Times’s own websites and mobile applications, rather

than exploit The Times’s content to keep users within their own search ecosystem. *Id.* ¶ 52. Defendants’ generative AI products threaten to divert that traffic. *Id.* ¶ 157.

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.” Compl. ¶ 56. OpenAI shed that nonprofit status in 2019, and is currently a full-blown commercial enterprise, with valuations as high as \$90 billion. *Id.* ¶¶ 57-58.

From 2018 through 2023, OpenAI developed a series of “large language models” or “LLMs,” 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. 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.* ¶ 83.

While 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.* ¶ 78. Training these models involved collecting and storing text content to create training datasets, and then processing that content through the GPT models. *Id.* ¶ 84. OpenAI worked with Microsoft to build these training datasets, which OpenAI does not dispute are packed with Times content. *Id.* ¶¶ 66-74. The exact number of Times works that Defendants copied to train their models is currently unknown, including because Defendants have not publicly disclosed the makeup of the datasets used to train GPT-3 and each subsequent model. *Id.* ¶¶ 58, 84.

Before filing suit, The Times sought to determine whether Defendants infringed its copyrights and uncovered overwhelming evidence that it had. The Complaint includes more than a hundred examples demonstrating that Defendants’ models “memorized” copies of Times works. *Id.* ¶¶ 98-101; Ex. J. The most updated model (GPT-4 LLM) outputs verbatim copies of significant portions of Times works and/or detailed summaries of those works when prompted. *Id.* That memorization is a product of how the models were trained. *Id.* ¶ 103. Defendants knew or should

have known they were infringing The Times’s copyrights, particularly because of The Times’s CMI on its content, publications, and websites. *Id.* ¶¶ 124-26. Not only did Defendants ignore this CMI, they designed the training process to remove it. *Id.*

C. Defendants’ user-facing products further infringe The Times’s copyrights.

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 ChatGPT. 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 paraphrase Times works retrieved and copied in response to user search queries in real time. *Id.* ¶ 102.

ChatGPT, a text-generating chatbot that mimics natural language in response to user prompts, initially produced only the first type of infringing outputs. *Id.* ¶ 61. After its November 2022 release, ChatGPT became an instant viral sensation, reaching one million users within a month and gaining over 100 million users within three months. *Id.* As shown in numerous examples throughout the Complaint, ChatGPT will display copies or derivatives of Times works memorized by the underlying models. *Id.* ¶¶ 103-07.

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.* ¶¶ 108, 118 Such “synthetic search” products 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.* Defendants know their products infringe on The Times’s copyrights. *Id.* ¶¶ 125-26.

Defendants benefited substantially from training on and using Times content without paying. Compl. ¶¶ 6, 92, 156. OpenAI’s 2024 revenues are projected to be over \$1 billion, *id.* ¶ 57, and Microsoft’s deployment of Times-trained models throughout its product line helped boost its market capitalization by a trillion dollars in the past year alone, *id.* ¶ 6. The Times has been harmed by Defendants’ free-riding. The Times limits the content it makes accessible for free and prohibits the use of its works for commercial uses absent a specific authorization. *Id.* ¶ 156. Yet Defendants have misappropriated almost a century’s worth of copyrighted content, without paying fair compensation. Compl. ¶ 157. Similarly, Defendants’ unlawful conduct threatens to divert readers away from accessing Times content directly from The Times, thus reducing the revenue that funds its journalism. *Id.* ¶ 157.

D. The Times files suit to protect its works.

In April 2023, The Times reached out to Defendants to inform them that their tools were infringing its copyrights. *Id.* ¶¶ 54, 126. Defendants persisted in their unlawful conduct, even while the parties engaged in negotiations through which The Times sought “to ensure it received fair value for the use of its content, facilitate the continuation of a healthy news ecosystem, and help develop GenAI technology in a responsible way that benefits society and supports a well-informed public.” *Id.* ¶ 7. 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 February 26, 2024, OpenAI filed this Motion, devoting ten pages of its brief to promoting Defendants’ products and another four pages to improperly attempting to refute facts

alleged in the Complaint. Ultimately, OpenAI seeks dismissal of (i) a narrow portion of the copyright infringement claim as untimely; (ii) the contributory infringement claim, (iii) the DMCA claim, and (iv) the unfair competition by misappropriation claim. The Times now opposes.

III. 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.* The Court “must limit itself to the facts stated in the complaint, documents attached to the complaint as exhibits, and documents incorporated by reference in the complaint.” *Grimes v. N.Y. & Presbyterian Hosp.*, 2024 WL 816208, at *4 (S.D.N.Y. Feb. 26, 2024).

IV. ARGUMENT

A. The Times’s direct infringement claim is timely.

OpenAI’s statute-of-limitations argument is narrow, addressing only the training of GPT-2 in 2019 and GPT-3 in 2020 and the construction of datasets used for those models. *See Mot.* at 15. It does not apply to the “orders of magnitude more powerful” GPT-3.5 and GPT-4 models developed in 2022 and 2023, respectively, nor does it relate to the recent deployment of Defendants’ user-facing products (e.g., ChatGPT and Copilot).¹ Compl. ¶¶ 61, 72.

But even OpenAI’s narrow argument fails because it ignores that, under the discovery rule, claims “do not accrue until the copyright holder discovers, or with due diligence should have discovered, the infringement.” *Sohm v. Scholastic Inc.*, 959 F.3d 39, 50 (2d Cir. 2020). OpenAI

¹ Nor could it apply to these models and products, since any claims based on events that occurred within the past two years plainly fall within the Copyright Act’s three-year statute of limitations. Compl. ¶ 83; 17 U.S.C. § 507(b). The argument also does not apply to any work that Defendants did on the GPT-2 and GPT-3 models after December 2020.

“bears the burden of proof” on this issue because “the statute of limitations [is] an affirmative defense.” *Hirsch v. Rehs Galleries, Inc.*, 2020 WL 917213, at *5 (S.D.N.Y. Feb. 26, 2020). OpenAI must “establish[] that [The Times] should be charged with constructive notice of the alleged infringement.” *McGlynn v. Sinovision Inc.*, 2024 WL 643021, at *2 (S.D.N.Y. Feb. 15, 2024).

That burden is a heavy one. The motion must be denied unless OpenAI proves that it is “clear from the face of the complaint” that “[The Times’s] claims are time-barred”—i.e., that by December 26, 2020, The Times should have been aware of the infringement. *Id.* At this early juncture, “even some doubt” necessitates denial. *PK Music Performance, Inc. v. Timberlake*, 2018 WL 4759737, at *7 (S.D.N.Y. Sept. 30, 2018). OpenAI’s entire argument on this point is relegated to a footnote that paraphrases the standard, cites nothing, and does not explain what allegedly put The Times on notice of its claims. Mot. at 15 n.33.

Without mentioning the discovery rule, the Background section of the brief raises factual arguments about when OpenAI “disclosed” the infringement. Mot. at 6-7. This portion of the brief reads like a summary judgment motion. Citing a handful of documents, OpenAI argues that, by 2020, it was “common knowledge” that some datasets used to train models “included numerous articles published by The Times.” Mot. at 6. But OpenAI admits that its own existence was not well-known until it released ChatGPT, in November 2022. *See* Mot. at 7; *see also* Compl. ¶ 61 (OpenAI did not become “a household name” until November 2022, when it released ChatGPT).

OpenAI’s identification of some public documents that provided information about the models’ training data sets (Mot. at 6 & n.19) is entirely insufficient to dismiss portions of The Times’s claim as untimely.² “[A] copyright holder does not have a general duty to ‘police the

² OpenAI’s reliance on these extraneous documents to raise factual disputes is improper. Even if their existence were judicially noticeable, OpenAI cannot rely on them “to prove the truth of their contents.” *Hesse v. Godiva Chocolatier, Inc.*, 463 F. Supp. 3d 453, 462 (S.D.N.Y. 2020) (cited by OpenAI, Mot. at 2 n.8). Courts refuse to take judicial notice

internet to discover” potentially infringing conduct, let alone scour and piece together fragments from academic articles. *Parisienne v. Scripps Media, Inc.*, 2021 WL 3668084, at *2 (S.D.N.Y. Aug. 17, 2021) (at the pleadings stage, rejecting the argument that the plaintiff “should have discovered the copyright infringement . . . when the article was posted on the website”). OpenAI does not even try to explain why The Times should have been aware, in 2020, of this small handful of documents. Whether and when these documents were available to The Times, what they disclose, and what could and should have been inferred from their contents are quintessential fact questions that cannot be resolved against The Times at this stage.

Even if The Times had known of the documents OpenAI cites, those documents do not remotely reveal the full scope of Defendants’ conduct in 2020. The “2019 snapshot of Common Crawl” (Mot. at 6 n.19) comes from a 2021 publication, making it irrelevant for this analysis since The Times filed this case in December 2023—within three years. *See* Compl. ¶ 88 & n.21. The GPT-2 and GPT-3 Papers (Mot. at 6 & n.19) did not even mention The Times, much less disclose that the training datasets included Times works. The “GPT-2 model card” cited by OpenAI (*id.*) represents that the models would “primarily” be “used by researchers,” which is a far cry from how they are used today. Relatedly, publication by The Times of news articles about OpenAI in 2020 does not establish The Times’s awareness of OpenAI’s copyright infringement, particularly when the articles emphasize that OpenAI only allowed a small number of people to access the models. *See* Mot. at 7-8 (citing two Times articles). At a bare minimum, there is “some doubt”

of documents on a motion to dismiss, where, as here, the proffering party is using the “submission to prove the substantive issue before the Court.” *Calcutti v. SBU, Inc.*, 273 F. Supp. 2d 488, 499 (S.D.N.Y. 2003). Similarly, the incorporation-by-reference doctrine does not cover documents—even those cited in a complaint—about which there is dispute regarding their authenticity or accuracy. *See Faulkner v. Beer*, 463 F.3d 130, 134 (2d Cir. 2006).

about when the claim accrued, and that doubt necessitates denial of the motion. *See PK Music Performance, Inc.*, 2018 WL 4759737, at *3.³

B. The Times states a contributory infringement claim.

OpenAI does not challenge The Times’s direct infringement claim, and instead seeks dismissal of The Times’s “alternative” contributory infringement claim. Compl. ¶ 179. “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). OpenAI argues only that The Times 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 Recs., LLC v. Doe 3*, 604 F.3d 110, 118 (2d Cir. 2010) (emphasis added); *ReDigi*, 934 F. Supp. 2d at 658. The Complaint sufficiently pleads OpenAI’s actual knowledge for at least five independent and legally sufficient reasons.

First, contrary to OpenAI’s assertion (Mot. at 16), The Times’s contributory infringement claim is not limited to “the Times’s creation of [the] outputs” cited in the Complaint, but instead extends to circumstances in which “an *end-user* may be liable as a direct infringer based on output of GPT-based products.” Compl. ¶ 179 (emphasis added). OpenAI presumably construes the claim in this narrow way because it believes that The Times was obligated to identify, at the pleading stage, every third party who has infringed specific Times articles by way of Defendants’ products.

³ OpenAI ignores that discovery into OpenAI’s construction of and use of the datasets for training GPT-2, GPT-3, and any other models before December 26, 2020, is entirely appropriate here insofar as those activities implicate any ongoing infringement occurring within the statute of limitations period. These ongoing activities may include, *inter alia*, OpenAI’s continued unauthorized reproduction and dissemination of its training datasets, the models, and any output from those models containing copies or derivatives of The Times’s copyrighted content.

OpenAI is wrong. See *Arista Recs. LLC v. Usenet.com, Inc.*, 633 F. Supp. 2d 124, 154 (S.D.N.Y. 2009) (“[K]nowledge of specific infringements is not required to support a finding of contributory infringement.”); *Arista Recs., Inc. v. Flea World, Inc.*, 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 Times 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’ service”). OpenAI’s citation to *Tiffany (NJ) Inc. v. eBay Inc.*, 600 F.3d 93 (2d Cir. 2010) (Mot. at 16 n.34) provides no help because that case applied the higher standard for contributory trademark infringement claims—a standard the *Tiffany* court expressly acknowledged is different from copyright law. *Id.* at 108.

Second, The Times has alleged that “Defendants were aware of many examples of copyright infringement after ChatGPT, Browse with Bing, and Bing Chat were released, some of which were highly publicized.” Compl. ¶ 126. In response, OpenAI asserts that “people do not use OpenAI’s products” to infringe, Mot. at 2, a dubious assertion given the “highly publicized” reports of infringement. Consistent with the allegations in the Complaint, public reports contradict OpenAI’s contention that its products have not been used to serve up paywall-protected content, underscoring the need for discovery.⁴

⁴ 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 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-bing-chatbot/>

It is reasonable to infer that OpenAI was aware of these reports of infringement, especially given the Complaint’s ample examples showing Defendants’ products generating infringing outputs in response to user queries. Compl. ¶¶ 104-22. And it is particularly inappropriate to make a finding about OpenAI’s knowledge at this stage because the best evidence of OpenAI’s knowledge is within its own possession. *See Argo Contracting Corp. v. Paint City Contractors, Inc.*, 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.*, 58 F. Supp. 2d 228, 251 (S.D.N.Y. 1999)). In short, OpenAI’s *ipse dixit* 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); *see also Rosen v. Amazon.com, Inc.*, 2014 WL 12597073, at *2 (C.D. Cal. May 28, 2014) (denying motion to dismiss contributory infringement claim because “the question of [defendant’s] knowledge, which is key to contributory liability, cannot be answered in the absence of admissible evidence.”).

Third, as alleged in the Complaint, OpenAI knew that its products were contributing to copyright infringement because The Times told it so. *See* Compl. ¶ 54 (alleging “The Times reached out to Microsoft and OpenAI in April 2023 to raise intellectual property concerns”), *id.* ¶ 126 (“[A]fter the release of ChatGPT and Bing Chat, The Times reached out to Defendants **to inform them that their tools infringed its copyrighted works.**” (emphasis added)). OpenAI ignores these allegations, claiming instead that The Times only alleges “generalized knowledge 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 REVIEW (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.’”).

possibility of infringement.” Mot. at 17. Its failure to acknowledge The Times’s outreach is particularly striking because OpenAI relies on a case explaining that “cease-and-desist letters” are “traditional indicia of actual or constructive knowledge” of contributory infringement. *Hartmann v. Popcornflix.com LLC*, 2023 WL 5715222, at *6 (S.D.N.Y. Sept. 5, 2023); Mot. at 16 n.35. “[I]f a computer system operator learns of specific infringing material available on his system [i.e., copyrighted works] and fails to purge such material from the system, the operator knows of and contributes to direct infringement.” *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1021 (9th Cir. 2001). This is precisely what happened here, where The Times informed OpenAI that its models were generating infringing outputs of Times works. *See ReDigi*, 934 F. Supp. 2d at 658 (reasoning that defendant “knew or should have known that its service would encourage infringement,” including based on a “cease-and-desist letter” that “advis[ed] [] defendant that its website violated . . . copyrights”); *Arista Recs. LLC.*, 633 F. Supp. 2d at 155 (“Defendants knew or should have known of infringement by its users,” including because “Defendants were explicitly put on notice of the existence of thousands of copies of Plaintiffs’ copyrighted sound recordings available on its service”).

Fourth, statements by OpenAI representatives support The Times’s claim of knowledge. 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. ¶ 124 (emphasis added). Such facts support The Times’s allegation that OpenAI knew it was contributing to copyright infringement. *See Arista Recs. LLC.*, 633 F. Supp. 2d at 155 (relying on “defendants’ employees’ own statements” to find that “defendants knew or should have known of infringement by its users”).

Fifth, OpenAI’s own concessions in the Motion show that it was well aware that, as pled in the Complaint, memorized training data was being generated by ChatGPT. *See* Compl. ¶ 18 & n.18 (citing Brown et al., Language Models are Few-Shot Learners (2020), <https://arxiv.org/pdf/2005.14165.pdf>, at 9 (acknowledging a “major methodological concern with language models pretrained on a broad swath of internet data, particularly large models with the capacity to memorize vast amounts of content”)). OpenAI admits as much by asserting that “[t]raining data regurgitation . . . is a *problem* that researchers at OpenAI and elsewhere work hard to address.” Mot. at 11. OpenAI could not have been “work[ing] hard” to address this “problem” without being aware of it. Moreover, OpenAI admits that it is tracking individual user interactions, including of The Times’s examples of memorization of copyrighted content reflected in Exhibit J to the Complaint (Mot. at 2), bolstering The Times’s allegation that OpenAI knew that The Times’s copyrighted works were being infringed.

OpenAI’s arguments to the contrary are based on a misapprehension of the law. OpenAI wrongly implies that constructive notice is insufficient for this claim, and that a plaintiff must also allege a defendant “took deliberate actions to avoid learning about the infringement.” Mot. at 16. Even if “constructive notice” were critical to this claim—which it is not, given The Times’s allegations of *actual* knowledge—OpenAI is wrong on the law because the standard permits liability for those who “‘know or have reason to know’ of the direct infringement.” *Doe 3*, 604 F.3d at 118; *see also Arista Recs., Inc. v. Mp3Board, Inc.*, 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” (emphasis added)). Even the cases cited by OpenAI

(Mot. at 16 n.35) show that constructive notice suffices, including recent decisions from other courts in this district.⁵

And OpenAI's reliance on *Viacom International, Inc. v. YouTube, Inc.*, 676 F.3d 19 (2d Cir. 2012), Mot. at 16 n.34, a case denying summary judgment because "emails request[ing] the identification and removal of" infringing footage created a material factual dispute, is perplexing given The Times's allegations regarding its infringement notices to Defendants.⁶

C. The Times states DMCA claims on multiple grounds.

The Times asserts claims under two subsections of the Digital Millennium Copyright Act ("DMCA"): Section 1202(b)(1), which proscribes the intentional removal or alteration of copyright management information ("CMI"), and Section 1202(b)(3), which proscribes the distribution of works while knowing that CMI has been removed or altered.

Defendants violated these provisions in two ways: (1) by "remov[ing] The Times's copyright management information in building the training datasets containing millions of copies of Times works," Compl. ¶ 184, in violation of § 1202(b)(1), and (2) by "remov[ing] The Times's [CMI]" in the process of "generating outputs" from Defendants' user-facing products like ChatGPT, *id.* ¶¶ 185-86, in violation of both § 1202(b)(1) and § 1202(b)(3). OpenAI challenges each of these theories on different grounds, including by (again) raising disputed fact issues.

⁵ *E.g.*, *Hartmann v. Apple, Inc.*, 2021 WL 4267820, at *7 (S.D.N.Y. Sept. 20, 2021) (plaintiff must allege defendant "knew or had reason to know"); *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").

⁶ OpenAI's remaining cases are out-of-circuit, and in any event unhelpful to its argument. In *Luvdarts, LLC v. AT & T Mobility, LLC*, 710 F.3d 1068, 1072 (9th Cir. 2013), the plaintiff relied exclusively on "notices" sent to the defendant-carriers, which "failed to notify the Carriers of any meaningful fact." The Times, by contrast, alleges that it put OpenAI on notice of the infringement, and that even OpenAI board members expressed concern. *BMG Rights Mgmt. (US) LLC v. Cox Communications, Inc.*, 881 F.3d 293, 308 (4th Cir. 2018), explained that "willful blindness" suffices for contributory infringement. As noted above, OpenAI had actual knowledge of infringement. But at a minimum, the allegations suggest willful blindness.

1. *The Times identifies the CMI at issue and alleges it is conveyed with its works.*

OpenAI’s argument that The Times does not “specify the CMI at issue” (Mot. at 17) is belied by its concession that the Complaint includes a “firm allegation” doing exactly that. *See* Mot. at 18 (citing Compl. ¶ 125, which explains how The Times places “copyright notices” and “link[s] to its terms of service” on “every page of its website,” and alleging that Defendants removed this CMI); *see also* Compl. ¶¶ 182, 187 (listing the at-issue CMI). This information meets the statutory definition of CMI. *See* 17 U.S.C § 1202(c). OpenAI’s identification of a single Times link that omits certain CMI (Mot. at 18) simply raises factual questions that cannot be answered now, including whether Defendants scraped the version of the page linked in the Motion or another version of the *same* article that contains The Times’s standard CMI.⁷ Nor does *Andersen v. Stability AI Ltd.*, 2023 WL 7132064, at *11 (N.D. Cal. Oct. 30, 2023), support dismissal, as there was “nothing in th[at] Complaint” to describe the CMI that was removed, making the allegation “wholly conclusory.” Not so here.

OpenAI also argues that The Times’s copyright notices and other CMI are too “small” and far down “at the bottom of the page” to meet the statutory definition of “conveyed.” Mot. at 18; *see also* 17 U.S.C. § 1202(c) (to qualify as CMI, the information must be “conveyed in connection with copies . . . of a work”). That argument reflects a misunderstanding of the term “conveyed,” which “is used in its broadest sense.” *Janik v. SMG Media, Inc.*, 2018 WL 345111, at *12 (S.D.N.Y. Jan. 10, 2018). The statute “merely requires that the information be accessible in conjunction with, or appear with, the work being accessed.” *Id.* Here, the CMI does in fact “appear with” the at-issue Times works, which is all that is required. *Id.*; *see also Pierson v. Infinity Music & Ent., Inc.*, 300 F. Supp. 3d 390, 396 (D. Conn. 2018) (terms of use were “conveyed in connection with” the at-

⁷ John Branch, *Snow Fall: The Avalanche at Tunnel Creek*, THE NEW YORK TIMES (Dec. 20, 2012), <https://www.nytimes.com/2012/12/21/sports/snow-fall-the-avalanche-at-tunnel-creek.html>.

issue works “because a link to the Terms of Use appeared at the bottom of each relevant webpage”). The specific “location of CMI” is at most relevant to OpenAI’s “intent,” which is a “fact issue [that] cannot be resolved on a motion to dismiss.” *Agence France Presse v. Morel*, 769 F. Supp. 2d 295, 305 (S.D.N.Y. 2011); *BanxCorp*, 723 F. Supp. 2d at 610-11 (similar).

2. *The Times states a DMCA claim under § 1202(b)(1).*

To state a claim under § 1202(b)(1), “[c]ourts have applied [the DMCA] in a straightforward manner such that Plaintiffs here need only allege (1) the existence of CMI . . . ; (2) removal and/or alteration of that information; and (3) that the removal and/or alteration was done intentionally.” *BanxCorp*, 723 F. Supp. 2d at 609.

a) The Times states a § 1202(b)(1) claim based on training.

OpenAI raises four arguments against this claim: that it is partially time-barred because the building of some training datasets occurred more than three years ago, that The Times fails to plausibly allege the removal of CMI, that The Times fails to plausibly allege the exclusion of CMI “by design,” and that The Times fails to plausibly allege that the removal of CMI could enable infringement. Mot. at 18-20. Each argument fails.

First, there is no statute-of-limitations problem for the reasons explained above. *See supra* § IV.A. OpenAI’s argument as applied to the DMCA claims is even weaker. Even if The Times were aware in 2020 that some training datasets included Times works, there was no way for The Times (or anyone) to discover whether CMI was removed during the training process since Defendants did not release a public-facing product until November 2022. Compl. ¶ 61.

Second, OpenAI argues that The Times “fails to plausibly allege that any CMI was removed” during the training process. Mot. at 18. Yet the Complaint includes over a hundred examples demonstrating how Defendants’ models output verbatim copies of Times works, and every one of those outputs omits CMI. Compl. Ex. J; *id.* ¶¶ 4-5, 98-122, 130. The omission of CMI

in the models' outputs suggests that CMI was removed during the training process; otherwise the models would have outputted all CMI as well. "Providing an actual example of the allegedly infringing [product] is obviously more than a conclusory allegation." *BanxCorp*, 723 F. Supp. 2d at 610 (denying motion to dismiss DMCA claim); *see also Devocean Jewelry LLC v. Associated Newspapers Ltd.*, 2016 WL 6135662, at *2 (S.D.N.Y. Oct. 19, 2016) (similar). OpenAI's so-called "particularly damning" argument—that "some CMI was preserved" (Mot. at 19)—also makes no sense because the statute proscribes removal of "any" CMI. 17 U.S.C. § 1202(b)(1). Defendants cannot evade liability by retaining "some CMI," like the "publication date" of an article (which the Complaint does not include in its list of removed CMI, Compl. ¶¶ 182, 187). Mot. at 19 n.40. Equally perplexing is OpenAI's assertion that "the Complaint lacks allegations about the inclusion" of Times CMI in third-party datasets like Common Crawl. Mot. at 19. Because Common Crawl is a "copy of the Internet" that includes Times 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. ¶¶ 88-89.

Third, OpenAI half-heartedly (in two sentences) attacks The Times's allegation that the training process was "designed" to remove CMI. Mot. at 19. At this stage, The Times's allegations "support a reasonable inference that Defendants intentionally designed the programs to remove CMI"; otherwise, the CMI would have appeared in the verbatim outputs. *Doe 1 v. GitHub, Inc.*, 2023 WL 3449131, at *11 (N.D. Cal. May 11, 2023) (denying motion to dismiss DMCA claims, alleging that defendant "trained [its] programs to ignore or remove CMI"). *Tremblay* does not alter this result, because it did not involve "an identical set of allegations" (Mot. at 18) – unlike The Times, those plaintiffs failed to include any allegations about infringing outputs. 2024 WL 557720, at *5 (N.D. Cal. Feb. 12, 2024).

Fourth, OpenAI argues that Defendants’ removal of CMI did not “enable[]” infringement by end-users of the products. Mot. at 20. But “nothing in the statutory language limits its applicability to such downstream infringement.” *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020). Here, removal of CMI “facilitates” or “conceals” OpenAI’s *own* infringement. Compl. ¶ 125. The DMCA’s so-called “double-scienter” requirement mandates that “the defendant who distributed improperly attributed copyrighted material [] have actual knowledge that CMI has been removed or altered without authority of the copyright owner or the law, as well as actual or constructive knowledge that such distribution will induce, enable, facilitate, or conceal an infringement.” *Mango*, 970 F.3d at 171 (cleaned up). And “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.” *Mango*, 970 F.3d at 172 (emphasis added). *Shihab v. Complex Media, Inc.*, 2022 WL 3544149, at *5 (S.D.N.Y. Aug. 17, 2022) (denying motion to dismiss DMCA claim where defendant “knew or should have known” that its removal of CMI “would conceal its contemporaneous copyright infringement”).

At this stage, “[c]ourts must be ‘lenient in allowing scienter issues’” related to the removal of CMI to survive motions to dismiss. *Hirsch v. CBS Broad. Inc.*, 2017 WL 3393845, at *8 (S.D.N.Y. Aug. 4, 2017) (denying motion to dismiss DMCA claim) (quoting *In re DDAVP Antitrust Litig.*, 585 F.3d 677, 693 (2d Cir. 2009)); *see also Agence France Presse*, 769 F. Supp. 2d at 305 (finding that arguments about “intent” would only raise a “fact issue [that] cannot be resolved on a motion to dismiss”).⁸ At a minimum, “it is at least as reasonable to infer that the CMI was altered to conceal or facilitate the alleged infringement as for any other reason.” *Devocean Jewelry LLC*, 2016 WL 6135662, at *2. Furthermore, OpenAI’s “continued” conduct after The Times informed

⁸ *Zuma Press, Inc. v. Getty Images (US), Inc.*, 845 F. App’x 54, 58 (2d Cir. 2021), cited by OpenAI, is inapposite because it was decided on summary judgment after discovery showed the defendant’s lack of knowledge. Mot. at 17.

OpenAI of this problem “constitutes knowledge sufficient to satisfy the DMCA’s second scienter requirement.” *Wright v. Miah*, 2023 WL 6219435, at *10 (E.D.N.Y. Sept. 7, 2023).

OpenAI’s cases on this issue are, again, inapposite. The *Tremblay* plaintiffs focused their claim on how the removal of CMI could *induce third parties* to infringe. *See Tremblay*, 2024 WL 557720, at *4, while The Times’s theory is that the removal of CMI *conceal and facilitates* OpenAI’s *own* infringement. *Roberts v. BroadwayHD LLC*, 518 F. Supp. 3d 719, 737 (S.D.N.Y. 2021), is even further afield because the DMCA claim in that case was brought under a different part of the statute—section 1202(a), which deals with the provision and distribution of false CMI, and which does not implicate the double-scienter requirement that OpenAI is challenging.

b) The Times states a § 1202(b)(1) claim based on outputs.

OpenAI argues this claim should be dismissed because “there was no CMI to remove from the relevant text” since the outputs cited in the Complaint “feature text from the middle of articles.” Mot. at 21. But the “relevant text” here refers to the articles that GPT-4 memorized, not the memorized output itself. And that original text is accompanied by CMI “on every page of [The Times’s] websites” and articles. Compl. ¶ 125. Whether the outputs are “wholesale copies of entire Times articles” (Mot. at 20) is irrelevant because DMCA claims may be based on the removal of CMI from “excerpts” of works. *See Planck LLC v. Particle Media, Inc.*, 2021 WL 5113045, at *6 (S.D.N.Y. Nov. 3, 2021). To the extent OpenAI’s argument is that the outputs are not sufficiently similar to Times works, that “argument is better reserved for summary judgment when there is a fully developed record.” *Pilla v. Gilat*, 2020 WL 1309086, at *12 (S.D.N.Y. Mar. 19, 2020).

OpenAI’s cases are, again, far afield. The allegedly infringing material in *Fischer* “b[ore] 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). As noted, the *Tremblay* plaintiffs did

not “provid[e] any indication as to what [the infringing] outputs entail,” *Tremblay*, 2024 WL 557720, at *5, which situates them very differently. OpenAI’s musings about how The Times’s theory would ascribe liability to journalists who include block quotes in a book review (Mot. at 21) ignores the fact-bound nature of the fair use doctrine. In due time, The Times will demonstrate why that defense does not apply to OpenAI’s conduct, but that issue is not yet before the Court.

3. ***The Times states a DMCA claim under § 1202(b)(3).***

OpenAI also argues that The Times failed to allege the “distribution” of outputs as required by § 1202(b)(3). According to OpenAI, the CMI-stripped content did not involve a “sale or transfer of ownership” but merely “public display” of that content. Mot. at 20. OpenAI’s argument overlooks how its products function. ChatGPT provides verbatim copies of Times works without CMI “in response to user prompts.” Compl. ¶ 103. 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 v. Plot Commerce*, 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)”). That is precisely what OpenAI is doing here.

4. ***The Times has standing to sue under the DMCA.***

OpenAI also asserts that The Times lacks standing because it does not allege that it was “injured by” Defendants’ DMCA violations. Mot. at 22. As explained above, OpenAI’s removal of CMI conceals and facilitates its own ongoing copyright infringement, which harms The Times in numerous ways. *See, e.g.*, Compl. ¶¶ 154-57. OpenAI’s focus on whether the outputs “reference[] The Times by name” (Mot. at 22) overlooks the fact that generating content from The Times without its CMI creates the false impression that OpenAI has the right to distribute this

work, thereby harming The Times’s licensing market since others will not pay for work that they can acquire for free. Defendants’ removal of CMI also “ma[kes] it easier for [] potential infringers” to violate The Times’s copyrights because it provides another method to misappropriate Times content. *Reilly*, 2016 WL 6837895, at *8. OpenAI’s cases are inapposite.⁹

Finally, even if it became difficult to quantify the precise harm arising from OpenAI’s DMCA violations (it is too soon to try) “courts have generally found an award [of statutory damages] appropriate as it provides fair compensation to the copyright holder and effectively deters future infringement.” *Olusola v. Don Coqui Holding Co., LLC*, 2021 WL 631031, at *5 (E.D.N.Y. Feb. 18, 2021). Statutory damages for DMCA violations are appropriate even “where there is little or no evidence to show that the DMCA violation increased the actual injury to the plaintiff.” *Miller v. Netventure24 LLC*, 2021 WL 3934262, at *8 (S.D.N.Y. Aug. 6, 2021).

D. The Times’s “hot-news” unfair competition by misappropriation claim is not preempted by the Copyright Act.

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-

⁹ *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999) (summary judgment opinion that did not even address whether the plaintiff was injured by a DMCA violation); *Steele v. Bongiovi*, 784 F. Supp. 2d 94, 98 (D. Mass. 2011) (dismissing DMCA claim where the alleged injury was that the defendants’ DMCA violations “caused [plaintiff] to lose” a prior copyright lawsuit).

1476 at 132, reprinted in 1976 U.S.C.C.A.N. at 5748 (footnote omitted), *quoted in Nat'l Basketball Ass'n v. Motorola, Inc.*, 105 F.3d 841, 850 (2d. Cir. 1997) (“*NBA*”).

OpenAI acknowledges that the Copyright Act does not preempt such a claim, but argues that it does not apply to these facts. *See* Mot. at 24 (recognizing “a ‘narrow’ preemption exception for ‘hot news’ claims endorsed by the Supreme Court in [*INS*]”); *Barclays Capital Inc. v. Theflyonthewall.com, Inc.*, 650 F.3d 876, 905-06 (2d Cir. 2011). OpenAI is wrong—if ever there was a case for the continued existence of this tort, it is this one, where Defendants have first trained their models to be able to speak like The Times and then have automated the use of those models to steal and rephrase Times content as their own.¹⁰ Compl. ¶ 72.

As explained in *Barclays*, 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 Associated Press and selling that news as though the defendant itself had gathered it.” *Barclays*, 650 F.3d at 902-03. The Second Circuit discussed two different formulations of a multi-part test articulated in *NBA* for asserting an *INS*-type non-preempted misappropriation claim. *Id.* at 898, 900.

Five-part test:

1. plaintiff generates or gathers information at a cost;
2. the information is time-sensitive;
3. defendant’s use of the information constitutes free-riding on plaintiff’s efforts;
4. the defendant is in direct competition with a product or service offered by the plaintiffs; and
5. the ability of other parties to free-ride on the efforts of the plaintiff or others would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.

¹⁰ Although OpenAI characterizes one part of The Times’s “hot-news” claim (confusingly called “The Text Claim”) as relating only to training, it is wrong. Mot. at 22-23. The Times brings this claim to assert misappropriation of both its breaking news content and product or service recommendations, and its allegations regarding training are merely intended to highlight the particular impropriety of using a Times-trained LLM to pilfer recent Times content.

Three-part test:

1. the time-sensitive value of factual information;
2. the free-riding by a defendant; and
3. the threat to the very existence of the product or service provided by the plaintiff.

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 (“Groundbreaking, In-Depth Journalism and Breaking News at Great Cost”); *id.* ¶ 193 (“time-sensitive breaking news” and “time-sensitive recommendations”). Defendants are free-riding by stealing The Times’s time-sensitive content generated at significant expense, and then publishing that content in direct competition with The Times through their products that leverage the Bing search index. *See, e.g., id.* ¶ 5 (“Defendants also 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.* ¶ 72 (describing synthetic search results that “purport to answer user queries directly”); *id.* ¶¶ 193-197 (Defendants are “free-riding on The Times’s significant efforts and investment of human capital to gather this information”). Finally, 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 generally id.* ¶¶ 108-123; *see also id.* ¶ 52.

These allegations are analogous to the hypothetical *INS*-type claim that the Second Circuit identified in *Barclays* as *not* being preempted: “[i]f a Firm were to collect and disseminate to some portion of the public 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.”); *Financial Info., Inc. v. Moody’s Investors Service, Inc.*, 808 F.2d 204, 209 (2d Cir. 1986) (“The ‘hot’ news doctrine is concerned with the copying and publication of information gathered by another before he has been able to utilize his competitive edge.”). Having misconstrued a portion of The Times’s claim as solely relating to training (Mot. at 22), OpenAI does not address how The Times’s actual claim for misappropriation of recent news stories fails to meet any part of the three- or five-part tests. The news article claim focuses on the diversion of revenue and traffic from synthetic search results that quote or paraphrase coverage of breaking news. Compl. ¶¶ 118-19.

With respect to the Wirecutter reviews, OpenAI argues only that the Complaint fails to meet the free-riding part because: (1) “OpenAI is not selling the Recommendations as its own”; and (2) “Wirecutter recommendations are not facts that The Times acquires through efforts akin to reporting.” Mot. at 25. Again, OpenAI bases its argument on factual disputes about the Complaint’s allegations, this time by arguing about the nature of Wirecutter’s reporting. *Id.*; see *DiBlasio v. Novello*, 344 F.3d 292, 304 (2d Cir. 2003) (holding that “a disputed issue of fact . . . is inappropriate to consider in the context of a Rule 12(b)(6) motion”). *Barclays* itself was decided after a full record was developed at trial. The Complaint adequately pleads facts, which taken as true, support the attributes of the free riding element for an *INS*-type claim.

Moreover, the “free riding” alleged is entirely distinguishable from that at issue in *Barclays*. There, 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 876, 903 (“It is Fly’s accurate attribution of the Recommendation to the creator that gives this news its value.”); *id.* at 905 (“[L]ike the defendants in *NBA* and unlike the defendant in *INS*, Fly has its own network and assembles and transmits data itself. . . . Fly’s employees are engaged in the financial industry equivalent of observing and summarizing facts about basketball games and selling those packaged facts to consumers.”).

By contrast, The Times has alleged that Defendants, by regurgitating time-sensitive Times content—sometimes with attribution to The Times and sometimes without—have misappropriated the underlying reporting and facts that The Times has invested significant resources into uncovering and publishing. By doing so, Defendants are disseminating The Times’s reporting “precisely at the point where the profit is to be reaped, in order to divert a material portion of the profit” from The Times to Defendants. *Barclays*, 650 F.3d at 904. Unlike here, in *Barclays* the fact that a brokerage firm had recommended a particular stock was itself the newsworthy event that could be reported on (by other humans). OpenAI 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. OpenAI also ignores entirely the allegation that it strips affiliate links from Wirecutter product reviews (Compl. ¶ 128), a fact that was not present in *Barclays* and one that underscores that this claim is distinct from The Times’s copyright infringement claims.

V. CONCLUSION

This Court should deny the motion in its entirety. Alternatively, the Court should grant leave to amend for any claim that is dismissed, particularly because The Times has not yet amended. *See Cortec Indus., Inc. v. Sum Holding L.P.*, 949 F.2d 42, 48 (2d Cir. 1991).

Dated: March 11, 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