

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

----- X
THE INTERCEPT MEDIA, INC., :
 : NO. 1:24-cv-01515-JSR
 :
 : Plaintiff, :
 :
 : v. :
 :
 :
 : OPENAI, INC., OPENAI GP, LLC, :
 : OPENAI, LLC, OPENAI OPKO LLC, :
 : OPENAI GLOBAL LLC, OAI :
 : CORPORATION, LLC, OPENAI :
 : HOLDINGS, LLC and MICROSOFT :
 : CORPORATION, X
 :
 : Defendants.

**MEMORANDUM OF LAW IN SUPPORT DEFENDANTS OPENAI, INC.,
OPENAI GP, LLC, OPENAI, LLC, OPENAI OPKO LLC, OPENAI GLOBAL LLC,
OAI CORPORATION, LLC, AND OPENAI HOLDINGS, LLC'S
MOTION TO DISMISS**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. PLAINTIFF’S ALLEGATIONS	2
III. LEGAL STANDARD.....	3
A. Motion to Dismiss for Lack of Subject Matter Jurisdiction Under Rule 12(b)(1)	3
B. Motion to Dismiss for Failure to State a Claim Under Rule 12(b)(6)	4
IV. ARGUMENT	4
A. Plaintiff Lacks Article III Standing.....	4
B. Plaintiff Fails to State a Claim Under Section 1202(b) of the DMCA	7
1. Copyright Management Information and the DMCA.....	7
2. Plaintiff Is Not Among Those Authorized to Sue Under 17 U.S.C. § 1203(a)	9
3. Plaintiff Fails to Specify the Works at Issue.....	10
4. Plaintiff’s Section 1202(b)(1) Claim Fails Because it Did Not Adequately Plead Scienter	12
5. Plaintiff’s Section 1202(b)(3) Claim Fails For Multiple Reasons	15
V. CONCLUSION.....	18

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Alan Ross Mach. Corp. v. Machinio Corp.</i> , No. 17-cv-3569, 2019 WL 1317664 (N.D. Ill. Mar. 22, 2019)	10, 14
<i>Ashcroft v. Iqbal</i> , 556 U.S. 662 (2009).....	4
<i>Associated Press v. All Headline News Corp.</i> , 608 F. Supp. 2d 454 (S.D.N.Y. 2009).....	8, 14
<i>Bell Atl. Corp. v. Twombly</i> , 550 U.S. 544 (2007).....	4
<i>Cable v. Agence Fr. Presse</i> , 728 F. Supp. 2d 977 (N.D. Ill. 2010)	8
<i>Carter v. HealthPort Techs., LLC</i> , 822 F.3d 47 (2d Cir. 2016).....	4, 7
<i>Cole v. John Wiley & Sons, Inc.</i> , No. 11-cv-2090, 2012 WL 3133520 (S.D.N.Y. Aug. 1, 2012).....	11
<i>Cortlandt St. Recovery Corp. v. Hellas Telecomms., S.a.r.l.</i> , 790 F.3d 411 (2d Cir. 2015).....	3
<i>Davis v. FEC</i> , 554 U.S. 724 (2008).....	4, 5
<i>Doe 1 v. GitHub, Inc.</i> , 672 F. Supp. 3d 837 (N.D. Cal. 2023)	5, 6
<i>Dow Jones & Co., Inc. v. Int’l Sec. Exch., Inc.</i> , 451 F.3d 295 (2d Cir. 2006).....	11, 12
<i>Dowd v. DeMarco</i> , 314 F. Supp. 3d 576 (S.D.N.Y. 2018).....	4
<i>Felix the Cat Prods., Inc. v. Cal. Clock Co.</i> , No. 04-cv-5714, 2007 WL 1032267 (S.D.N.Y. Mar. 30, 2007).....	11
<i>First Nationwide Bank v. Gelt Funding Corp.</i> , 27 F.3d 763 (2d Cir. 1994).....	13

Fischer v. Forrest,
286 F. Supp. 3d 590 (S.D.N.Y. 2018).....14, 16

Free Speech Sys., LLC v. Menzel,
390 F. Supp. 3d 1162 (N.D. Cal. 2019)12

Frost-Tsuji Architects v. Highway Inn, Inc.,
No. 13-cv-00496, 2015 WL 263556 (D. Haw. Jan. 21, 2015).....17

Joint Stock Co. Channel One Russ. Worldwide v. Infomir LLC,
No. 16-cv-1318, 2017 WL 696126 (S.D.N.Y. Feb. 15, 2017)11

Kelly v. Arriba Soft Corp.,
77 F. Supp. 2d 1116 (C.D. Cal. 1999), *rev'd in part on other grounds*,
336 F.3d 811 (9th Cir. 2003)9, 15

Lance v. Coffman,
549 U.S. 437 (2007).....3

Lujan v. Defs. of Wildlife,
504 U.S. 555 (1992).....4, 6

Mango v. BuzzFeed, Inc.,
970 F.3d 167 (2d Cir. 2020).....13

McGucken v. Shutterstock, Inc.,
No. 22-cv-00905, 2023 WL 6390530 (S.D.N.Y. Oct. 2, 2023).....13

Palmer Kane LLC v. Scholastic Corp.,
No. 12-cv-3890, 2013 WL 709276 (S.D.N.Y. Feb. 27, 2013)10

Reiffer v. NYC Luxury Limousine Ltd.,
No. 22-cv-2374, 2023 WL 4029400 (S.D.N.Y. June 15, 2023)9

Robert L. Stark Enters., Inc. v. Neptune Design Grp., LLC,
No. 16-cv-264, 2017 WL 1345195 (N.D. Ohio Apr. 12, 2017)17

Roberts v. BroadwayHD LLC,
518 F. Supp. 3d 719 (S.D.N.Y. 2021).....17

Sherwood 48 Assocs. v. Sony Corp. of Am.,
76 F. App'x 389 (2d Cir. 2003)10

Sitnet LLC v. Meta Platforms, Inc.,
No. 23-cv-6389, 2023 WL 6938283 (S.D.N.Y. Oct. 20, 2023).....10

Steele v. Bongiovi,
784 F. Supp. 2d 94 (D. Mass. 2011)9

Stevens v. CoreLogic,
899 F.3d 666 (9th Cir. 2018)15, 17

Tremblay v. OpenAI Inc.,
No. 23-cv-03416, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024).....13

Victor Elias Photography, LLC v. Ice Portal, Inc.,
43 F.4th 1313 (11th Cir. 2022)14

Wolo Mfg. Corp. v. ABC Corp.,
349 F. Supp. 3d 176 (E.D.N.Y. 2018)11

Zuma Press, Inc. v. Getty Images (US), Inc.,
845 F. App’x. 54 (2d Cir. 2021)9

Statutes

15 U.S.C. §§ 1051 et seq. (“Lanham Act”)10

17 U.S.C. § 508.....12

17 U.S.C. §§ 1201-1205 (Digital Millennium Copyright Act (“DMCA”))..... *passim*

35 U.S.C. §§ 1 et seq. (“Patent Act”).....10

Other Authorities

U.S. CONST. art. III3, 4, 6, 9

Fed. R. Civ. P.
8.....4
12(b)(1)3
12(b)(6)4

S. Rep. 105-190 (1998).....8

Julie E. Cohen, A Right To Read Anonymously: A Closer Look at “Copyright Management” in Cyberspace, 28 CONN. L. REV. 981 (1996).....7

U.S. Dep’t of Commerce, *Information Infrastructure Task Force, Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights* (1995).....8

I. INTRODUCTION

In recent years, rapid advances in the field of artificial intelligence (AI) have raised critical legal questions, including about the application of longstanding copyright principles to new technology. These questions have led to a series of lawsuits that raise genuinely important copyright issues ripe for judicial resolution. This case, however, is not one of them. Instead, it is a follow-on lawsuit against OpenAI that represents an attempt to join the fray, but without alleging the facts necessary to present the issues that Plaintiff wants to bring before this Court.

OpenAI is a pioneer in the field of AI and the creator of an AI service known as ChatGPT. In its current form, the technology allows users to submit text prompts and receive back content generated by the software and servers that comprise the ChatGPT service. That service, in turn, works by combining an underlying engine known as a “large language model” with additional measures intended to ensure the accuracy, appropriateness, safety, and utility of the outputs presented to users. The large language model is a type of “neural network.” It consists of a staggeringly large series of statistical correlations that yield rules about the constitutive elements of human language—correlations and rules the model “learned” by having been shown many, many examples of text. For the purpose of “training” a model of this type, it is the volume of text used, more than any particular selection of text, that really matters, for the simple reason that a truly massive quantity of samples is required to derive comprehensively accurate statistical representations of the concepts of grammar, semantics, and quirks that underlie ordinary human language.

Notably, this case does not involve a copyright infringement claim. Instead, the essence of Plaintiff’s complaint is that OpenAI allegedly removed author, title, and terms of use information from copies of Plaintiff’s unspecified “works of journalism” that OpenAI allegedly

used to “train” its large language models. (Compl. (ECF No. 1) ¶ 53.) To support its claims, Plaintiff avers generally that ChatGPT has “regurgitate[d]” or “mimic[ked]” copyrighted journalism in response to user prompts. (*Id.* ¶¶ 35-36.) But Plaintiff never offers a single instance in which ChatGPT actually is alleged to have done so. The complaint does not describe *any* ChatGPT output of *any* specific work of journalism, let alone one of Plaintiff’s. Indeed, the complaint fails even to identify Plaintiff’s “works of journalism” at all. The complaint thus suffers from both jurisdictional defects requiring dismissal of the entire complaint and particular defects in each pleaded cause of action under 17 U.S.C. § 1202(b).

For these reasons and the others articulated below, OpenAI respectfully seeks dismissal of the complaint.

II. PLAINTIFF’S ALLEGATIONS

Plaintiff is a news organization that publishes “copyrighted works of journalism” “on the internet.” (Compl. ¶¶ 8, 32.) It brings claims against seven OpenAI entities (which it refers to collectively as “the OpenAI defendants,” without differentiating among them) for violations of Section 1202(b) of the Digital Millennium Copyright Act. Plaintiff does not bring any claim for copyright infringement, nor provide any example of any of its articles that it believes any OpenAI entity has reproduced or distributed without its permission.

Instead, the core of Plaintiff’s allegations is that OpenAI developed its “ChatGPT AI products” by using “training sets” that included unspecified “works of journalism” from which OpenAI “intentionally removed author, title, copyright notice, and terms of use information” in an unspecified way. (Compl. ¶¶ 8, 38-43.) To support these claims, Plaintiff also contends that “[a]t least some of the time, ChatGPT provides or has provided responses to users that regurgitate verbatim or nearly verbatim copyright-protected works of journalism” or “that mimic

significant amounts of material from copyright-protected works of journalism.” (*Id.* ¶¶ 35-36.) But Plaintiff does not specifically identify a single “work of journalism” that it owns—let alone one that was supposedly used by OpenAI for training or contained in a ChatGPT output. Instead, it just says that the works at issue were ones “published on the internet.” (*Id.* ¶ 32.)

Plaintiff also alleges that “[e]arlier versions of ChatGPT (prior to GPT-4) were trained” on three datasets known as WebText, WebText2, and Common Crawl. (*Id.* ¶ 30.) Plaintiff alleges that WebText and WebText2 were created by the OpenAI defendants, while “Common Crawl originated elsewhere,” and that these datasets consist of “collections of links posted on the website Reddit” and “a scrape of most of the internet.” (*Id.* ¶ 31.) Plaintiff, however, does not allege whether any of its works actually appeared in these datasets. Instead, Plaintiff alleges only that “[t]housands” of unidentified examples of its works appeared in “recreated approximations” of the datasets. (*Id.* ¶ 38.) But Plaintiff offers no further information about who created these “approximations,” how they were created, or which of Plaintiff’s articles the “approximations” contained.

III. LEGAL STANDARD

A. Motion to Dismiss for Lack of Subject Matter Jurisdiction Under Rule 12(b)(1)

Article III “limits the jurisdiction of federal courts to ‘Cases’ and ‘Controversies.’” *Lance v. Coffman*, 549 U.S. 437, 439 (2007). If a plaintiff lacks Article III standing to bring a suit, the federal court lacks subject matter jurisdiction, and the suit must be dismissed under Rule 12(b)(1). *Cortlandt St. Recovery Corp. v. Hellas Telecomms., S.a.r.l.*, 790 F.3d 411, 416-17 (2d Cir. 2015). In evaluating a motion to dismiss for lack of standing, the “task of the district court is to determine whether the Pleading alleges facts that affirmatively and plausibly suggest that

the plaintiff has standing to sue.” *Carter v. HealthPort Techs., LLC*, 822 F.3d 47, 56 (2d Cir. 2016) (cleaned up).

B. Motion to Dismiss for Failure to State a Claim Under Rule 12(b)(6)

To satisfy Rule 8 and survive a Rule 12(b)(6) motion to dismiss, “a complaint must contain sufficient factual matter, accepted as true, to state a claim to relief that is plausible on its face.” *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (cleaned up). Moreover, “[f]actual allegations must be enough to raise a right to relief above the speculative level.” *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 555 (2007). “[A] formulaic recitation of the elements of a cause of action will not do.” *Id.* Further, “conclusory allegations or legal conclusions masquerading as factual conclusions will not suffice to prevent a motion to dismiss.” *Dowd v. DeMarco*, 314 F. Supp. 3d 576, 582 (S.D.N.Y. 2018) (quoting *Smith v. Local 819 I.B.T. Pension Plan*, 291 F.3d 236, 240 (2d Cir. 2002)).

IV. ARGUMENT

A. Plaintiff Lacks Article III Standing.

In order to bring a claim, a plaintiff must plead a cognizable injury to itself, not to someone else. Plaintiff alleges that ChatGPT has provided responses to users containing material from someone else’s “works of journalism.” (Compl. ¶¶ 35-37.) But Plaintiff has not pointed to a single example of a ChatGPT output that looks anything like one of its *own* copyrighted works. Plaintiff therefore has failed to plead that it suffered a cognizable injury under Article III of the U.S. Constitution. *See Lujan v. Defs. of Wildlife*, 504 U.S. 555, 560 (1992). The complaint should be dismissed for that reason alone.

In any lawsuit, the plaintiff bears the burden of establishing standing “for each claim [s]he seeks to press and for each form of relief that is sought.” *Davis v. FEC*, 554 U.S. 724, 734

(2008) (cleaned up). A plaintiff must allege: (1) an injury in fact that is concrete and particularized, as well as actual or imminent; (2) that the injury is fairly traceable to the challenged action; and (3) that it is likely (not merely speculative) that the injury will be redressed by a favorable decision. *Id.* at 733. Further, as another court recognized in a similar AI training case, an “increased risk of future harm alone is not sufficiently concrete to confer standing for damages.” *Doe 1 v. GitHub, Inc.*, 672 F. Supp. 3d 837, 850-51 (N.D. Cal. 2023) (citing *TransUnion LLC v. Ramirez*, 594 U.S. 413, 436 (2021)).

Here, the complaint contains no express allegation or explanation of whether and how Plaintiff was harmed. Plaintiff alleges that its journalism is “the result of significant investments by Plaintiff in the human and other resources necessary to report on the news.” (Compl. ¶ 33.) But even so, Plaintiff has not plausibly alleged that any ChatGPT output even resembled one of Plaintiff’s articles (let alone was identical to that article, as required by the DMCA, *see* §IV.B.5., *infra*). Nor has Plaintiff offered any plausible allegation of harm to its investments from the mere presence of some unidentified subset of articles as a miniscule fraction of the “scape of most of the internet” used to train models used for ChatGPT, in the absence of such an output. (*Id.* ¶ 30.)

Instead, the complaint makes generalized allegations that ChatGPT has outputted *someone’s* copyrighted works of journalism to users. (*See, e.g.*, Compl. ¶¶ 35-36.) But Plaintiff’s allegations do not plausibly suggest that *Plaintiff’s* articles (as opposed to some other new organization’s articles) were actually outputted by ChatGPT. Plaintiff *does not identify a single example* of ChatGPT outputting its “works of journalism,” or even anything substantially similar to or derived from its “works of journalism.” In fact, the complaint fails to identify with any reasonable specificity what its “works of journalism” *are*. It states only that those works are

“published on the internet.” (*Id.* ¶ 32.) And without information about what, exactly, plaintiff’s works are and whether ChatGPT has ever outputted them to a user, the complaint’s allegations do not state an injury that “affect[s] the plaintiff in a personal and individual way.” *Lujan*, 504 U.S. at 560 n.1. Accordingly, any injury connected to ChatGPT outputs is insufficiently particularized for Article III standing.

Doe 1 v. GitHub, Inc., another DMCA case brought against OpenAI, is instructive. In that case, an anonymous group of software developers alleged that OpenAI violated the DMCA by using their copyrighted computer code as training data. Similarly, there, plaintiffs never alleged that *their own copyrighted works* appeared in any output of the AI model. Instead, the software developers’ complaint merely identified “several instances in which [the AI model’s] output matched . . . code written by a Github user” who was not a plaintiff. *Doe 1*, 672 F. Supp. 3d at 850. The court found those allegations insufficient to confer standing for damages. *Id.* The Court also rejected the argument that including the developers’ code as training data created an “increased and imminent risk” that such injury could occur “at any time.” *Id.* Because the developers failed to “allege any additional, concrete harm associated with this increased risk of misappropriation,” the Court held such allegations “cannot provide standing for Plaintiffs’ damages claims.” *Id.* at 850-51. Thus, the Court dismissed all damages claims for lack of standing.

Here, as in *Doe 1*, Plaintiff does “not allege that [it] ha[s] suffered the injury [it] describe[s].” *Doe 1*, 672 F. Supp. 3d at 850. Just as the plaintiffs in *Doe 1* anchored their claim on an allegation that AI output has “matched licensed code written by [other] Github user[s],” *id.*, here too Plaintiff relies only on the general allegation that “[a]t least some of the time, ChatGPT provides . . . responses to users that regurgitate” or “mimic significant amounts of

material from copyright-protected works of journalism.” (Compl. ¶¶ 35-36.) The complaint fails to allege any examples of “regurgitat[ion]” or “mimic[ing]” happening to *Plaintiff*, nor any reason why it would happen to Plaintiff. Accordingly, the complaint fails to allege “facts that affirmatively and plausibly suggest that the plaintiff has standing to sue,” *Carter*, 822 F.3d at 56 (cleaned up), and Plaintiff’s claims should be dismissed.

B. Plaintiff Fails to State a Claim Under Section 1202(b) of the DMCA.

Plaintiff brings two claims against OpenAI under Section 1202(b) of the DMCA. As relevant here, Section 1202(b) prohibits the “intentional[] remov[al] or alter[ation]” of copyright management information (“CMI”), *see* 17 U.S.C. § 1202(b)(1), and the “distribut[ion]” of “copies of works . . . knowing that [CMI] has been removed or altered,” with “reasonable grounds to know” that the removal or distribution “will induce, enable, facilitate, or conceal” copyright infringement, *see id.* § 1202(b)(3). But Plaintiff’s claims are far afield from the statutory language and purpose of Section 1202, which is briefly surveyed below. That background illuminates the flaws in these claims. As addressed below, both of Plaintiff’s Section 1202(b) claims fail because Plaintiff (1) lacks statutory standing, (2) has not identified the works from which OpenAI allegedly removed CMI, and (3) has failed to adequately plead that OpenAI had reason to know its purported CMI removal and distribution of copies would induce, enable, facilitate, or conceal copyright infringement.

1. Copyright Management Information and the DMCA

Congress enacted Section 1202 of the DMCA as part of a policy initiative to “help copyright owners police their copyrights, in light of the otherwise trivial ease of generating and distributing unauthorized copies of their works throughout cyberspace.” Julie E. Cohen, *A Right To Read Anonymously: A Closer Look at “Copyright Management” in Cyberspace*, 28 CONN. L.

REV. 981, 990 (1996); *see also* S. Rep. 105-190 at 8, 11 n.18 (1998) (noting the overarching legislative purpose to “discourage piracy” on the Internet); U.S. Dep’t of Commerce, Information Infrastructure Task Force, *Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights* 235–36 (1995) (initially proposing §1202). To permit copyright owners to “track[] and monitor[]” how their works are used online, Section 1202 encourages owners to affix CMI to their works, and imposes penalties on defendants who frustrate those efforts by removing or altering that information. S. Rep. 105-190 at 16–17. Thus, in the typical Section 1202 case, the plaintiff might allege that it published, for example, a photograph along with a “photo credit and copyright notice,” and that the defendant both copied that photograph and “deliberately removed” the accompanying CMI—*e.g.*, by cropping out the photo credit—to hide its wrongdoing. *Cable v. Agence Fr. Presse*, 728 F. Supp. 2d 977, 978 (N.D. Ill. 2010); *see also Associated Press v. All Headline News Corp.*, 608 F. Supp. 2d 454, 458 (S.D.N.Y. 2009) (Section 1202 claim based on allegation that editors “instructed reporters to remove or alter the identification of the AP as author or copyright holder of [news] articles”).

To ensure the statute hewed closely to that purpose, Congress included in Section 1202 a “double-scienter” requirement. Simply removing CMI from a copyrighted work does not result in liability. A Section 1202 plaintiff must also plead and prove both (1) that the removal or alteration was done “intentionally” (or, for distribution-based claims, was done “knowing that [CMI] has been removed or altered”), and (2) that these acts were performed with “reasonable grounds to know . . . that [they] will induce, enable, facilitate, or conceal an infringement [of copyright].” 17 U.S.C. § 1202(b).

In the typical Section 1202 case, these elements are not difficult to plead. *See, e.g., Reiffer v. NYC Luxury Limousine Ltd.*, No. 22-cv-2374, 2023 WL 4029400, at *8 (S.D.N.Y. June 15, 2023) (scienter requirement satisfied because “Defendant saw Plaintiff’s attribution on the Work and removed it before uploading it to its website”). But the “double-scienter” requirement precludes liability in many circumstances, such as, for example, when omission of CMI occurs as an “unintended side effect” of a technological process—like scraping images from the internet without also scraping associated CMI. *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999), *rev’d in part on other grounds*, 336 F.3d 811 (9th Cir. 2003) (no Section 1202 liability arising from the fact that search engine “crawler did not include [CMI] when it indexed the images”); *see also Zuma Press, Inc. v. Getty Images (US), Inc.*, 845 F. App’x. 54, 58 (2d Cir. 2021) (finding no Section 1202(b)(3) liability where “purported changes to Plaintiffs’ CMI resulted not from some intentional act of which [defendant] was aware but from aberrations and mistakes in the automatic migration process itself, during which [defendant] processed approximately 7,000,000 images”). In other words, the incidental removal of CMI does not raise an inference that the defendant knew its actions would conceal infringement.

With that background in mind, OpenAI now turns to the substance of Plaintiff’s DMCA claims.

2. Plaintiff Is Not Among Those Authorized to Sue Under 17 U.S.C. § 1203(a).

Even assuming Plaintiff had standing under Article III—it does not—Plaintiff’s Section 1202(b) claims separately fail because Plaintiff is not within the class of plaintiffs that Congress authorized to sue under that section. “[T]o have standing” to sue for a DMCA violation, Plaintiff “must show that [it] was injured by *that* violation.” *Steele v. Bongiovi*, 784 F. Supp. 2d 94,

97-98 (D. Mass. 2011) (emphasis added); *see also* 17 U.S.C. § 1203(a) (“Any person injured by a violation of section 1201 or 1202 may bring a civil action”). Here, however, Plaintiff does not explain how it was injured by OpenAI’s alleged removal of CMI from its training dataset or its alleged sharing of that data with Microsoft. Even assuming Plaintiff’s works were included in OpenAI’s training dataset—which Plaintiff has not plausibly alleged, *see* §IV.B.4., *infra*—Plaintiff still has not alleged how a *lack of CMI* in an internal training dataset could be connected to any harm flowing from ChatGPT’s outputs (which, for that matter, it also has not plausibly alleged). This “failure to allege an injury” connected to the alleged DMCA violation is separately “fatal” to its claims. *See Alan Ross Mach. Corp. v. Machinio Corp.*, No. 17-cv-3569, 2019 WL 1317664, at *4 (N.D. Ill. Mar. 22, 2019) (dismissing DMCA claim for lack of standing under 17 U.S.C. § 1203(a)).

3. Plaintiff Fails to Specify the Works at Issue.

Plaintiff’s Section 1202 claims are also subject to dismissal because the complaint has not identified any works from which CMI was allegedly removed. As a preliminary matter, intellectual property claims of all stripes require a plaintiff “to specify which works are at issue.” *Palmer Kane LLC v. Scholastic Corp.*, No. 12-cv-3890, 2013 WL 709276, at *3 (S.D.N.Y. Feb. 27, 2013) (dismissing copyright infringement claim); *see also Sherwood 48 Assocs. v. Sony Corp. of Am.*, 76 F. App’x 389, 391 (2d Cir. 2003) (affirming dismissal of Lanham Act claim where plaintiffs “have not identified their purportedly protectable trade dress with precision”); *see also Sitnet LLC v. Meta Platforms, Inc.*, No. 23-cv-6389, 2023 WL 6938283, at *1 (S.D.N.Y. Oct. 20, 2023) (requiring patent infringement disclosures that identify “each accused apparatus, product, device, process, method, act, or other instrumentality” and demanding that the “identification shall be as specific as possible”). This principle does not depend on whether the

claim arises under the Copyright Act, the DMCA, the Lanham Act, or the Patent Act. Rather, it is because the “mere assertion” of infringement “without any factual allegations concerning the nature” of the infringement “does not give the defendants fair notice of the claims against them.” *Dow Jones & Co., Inc. v. Int’l Sec. Exch., Inc.*, 451 F.3d 295, 307 (2d Cir. 2006); *see also Felix the Cat Prods., Inc. v. Cal. Clock Co.*, No. 04-cv-5714, 2007 WL 1032267, at *4 (S.D.N.Y. Mar. 30, 2007) (dismissing copyright and trademark claims where plaintiff failed to identify specific works at issue and citing *Dow Jones*, 451 F.3d at 307).

Even though Plaintiff does not bring a copyright infringement claim here, copyright pleading standards are instructive as to what is required to “give [OpenAI] fair notice of the claims against [it].” *See Dow Jones*, 451 F.3d at 307. In copyright cases, a plaintiff may not base its claims “on overly-inclusive lists of copyrighted works” as a “substitute for allegations that specify the original works that are the subject of a copyright claim.” *Cole v. John Wiley & Sons, Inc.*, No. 11-cv-2090, 2012 WL 3133520, at *12 (S.D.N.Y. Aug. 1, 2012) (dismissing copyright infringement claim). Even where a complaint alleges “wholesale infringement of a large number of copyrighted works,” the plaintiff “must still identify, at a minimum, representative examples of the works allegedly infringed.” *Joint Stock Co. Channel One Russ. Worldwide v. Infomir LLC*, No. 16-cv-1318, 2017 WL 696126, at *14 (S.D.N.Y. Feb. 15, 2017). Applying that rule, courts in this district routinely reject attempts to plead copyright claims using open ended catch-all descriptors to identify the works at issue. *See, e.g., Wolo Mfg. Corp. v. ABC Corp.*, 349 F. Supp. 3d 176, 202 (E.D.N.Y. 2018) (dismissing copyright claims with respect to “unspecified ‘Other Works’”); *Cole*, 2012 WL 3133520, at *12-13 (dismissing complaint based on “vague and expansive allegations regarding which works are the subject of Plaintiff’s claims”). Accordingly, DMCA claims have also failed where a plaintiff “merely alleged that his

photographs ‘were altered to remove certain of [his] copyright management information’ without providing any facts to identify *which* photographs had CMI removed.” *Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1175 (N.D. Cal. 2019) (emphasis added).

Here, Plaintiff makes no attempt whatsoever to identify the works on which it predicates its Section 1202(b) claims. Plaintiff makes general allegations about its “works of journalism” that were “published on the internet.” (Compl. ¶ 32.) But this is too imprecise to “give [OpenAI] fair notice of the claims against [it].” *See Dow Jones*, 451 F.3d at 307. Indeed, Plaintiff’s other allegations serve only to highlight this imprecision—according to the complaint, “thousands of Plaintiff’s copyrighted works were included in Defendants’ training sets,” but Plaintiff does not (or cannot) identify a single one of them. (*See* Compl. ¶ 42.)

Even when Plaintiff filed its “Report on the Filing or Determination of an Action or Appeal Regarding a Copyright” on this Court’s docket, it still failed to identify the works at issue. (*See* ECF No. 8.) This Report—known as form AO 121—provides information to the Register of Copyrights about the works at issue in a copyright litigation, and must be filed within one month of bringing an action. 17 U.S.C. § 508. But on its form, Plaintiff only lists “Various” in the fields for “Title of Work” and “Author of Work.” (ECF No. 8.) As a result, not only does OpenAI lack fair notice of the claims against it, but the Register of Copyrights lacks the information required by 17 U.S.C. § 508. Without identifying specific works from which CMI was removed, Plaintiff fails to state a Section 1202(b) claim.

4. Plaintiff’s Section 1202(b)(1) Claim Fails Because it Did Not Adequately Plead Scierter.

Plaintiff also has not pled facts sufficient to meet the DMCA’s “double-scierter” requirement. To plead a violation of Section 1202(b)(1), Plaintiff must allege both that OpenAI intentionally removed CMI, and that OpenAI had “reasonable grounds to know” that such

removal would “induce, enable, facilitate, or conceal [copyright] infringement.” 17 U.S.C. § 1202(b); *see Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020). Further, Plaintiff must allege facts “from which one can infer that future infringement is likely, albeit not certain, to occur as a result of the removal” of CMI. *McGucken v. Shutterstock, Inc.*, No. 22-cv-00905, 2023 WL 6390530, at *11 (S.D.N.Y. Oct. 2, 2023) (quoting *Stevens v. CoreLogic*, 899 F.3d 666, 676 (9th Cir. 2018)). Plaintiff’s allegations, however, miss the mark.

As an initial matter, Plaintiff fails to plausibly allege the intentional removal CMI from any of its works. Plaintiff alleges that “[a]t least some of the time,” ChatGPT produces outputs that “regurgitate” or “mimic significant amounts of material from copyright-protected works of journalism without providing” the accompanying CMI. (Compl. ¶¶ 35-36.) From that, Plaintiff speculates that “ChatGPT would have learned to communicate [CMI] when providing responses to users unless Defendants trained it otherwise.” (*Id.* ¶ 39.) But this is a leap of logic that requires at least two “unwarranted deductions of fact,” *First Nationwide Bank v. Gelt Funding Corp.*, 27 F.3d 763, 771 (2d Cir. 1994): that Plaintiff’s unidentified works of journalism actually appeared in the “regurgitate[d]” or “mimic[ked]” outputs (Compl. ¶¶ 35-36), and that ChatGPT would have outputted the CMI but for an intentional choice by OpenAI to strip it from training data. In fact, this argument depends on yet a further unstated assumption: that the training process does not preserve CMI by design. Courts, however, have rejected this exact allegation as insufficient even when it was expressly included in the complaint. *See, e.g., Tremblay v. OpenAI Inc.*, No. 23-cv-03416, 2024 WL 557720, at *4 (N.D. Cal. Feb. 12, 2024) (allegation that “training process does not preserve any CMI” “by design” was “conclusory”).

Separately, even if OpenAI’s training process did result in the omission of CMI from the alleged copies in its training dataset, Plaintiff does not plausibly allege how that omission could

“induce, enable, facilitate, or conceal” the alleged copyright infringement, much less how OpenAI could have had “reason to know” that it would do so. 17 U.S.C. § 1202(b). “The point of CMI is to inform the *public* that something is copyrighted and to prevent infringement.” *Alan Ross Mach.*, 2019 WL 1317664 at *2 (emphasis added); *see also Fischer v. Forrest*, 286 F. Supp. 3d 590, 610-11 (S.D.N.Y. 2018) (“CMI exists to inform the public that a work is copyrighted and by whom The DMCA exists, in part, to protect that notice.”). But Plaintiff admits that OpenAI’s training dataset is not publicly accessible, (Compl. ¶ 29), so omission in the training dataset alone, even were that allegation proven, could not deprive the public of copyright management information.

In the typical Section 1202 case, a plaintiff might plead scienter by alleging, for example, that the defendant “instructed reporters to remove or alter the identification of the [original publisher]” before “distribut[ing] its articles to paying clients” who, as a result of the CMI removal, would have no way to know that the defendant had infringed another publisher’s copyright. *All Headline News*, 608 F. Supp. 2d at 458. In other words, there is a direct and obvious causal connection between the CMI omission and the “facilitat[ion]” of “infringement” to satisfy the statute’s double scienter requirement. 17 U.S.C. § 1202(b); *see also Victor Elias Photography, LLC v. Ice Portal, Inc.*, 43 F.4th 1313, 1325 (11th Cir. 2022) (“[T]he statute’s plain language requires some identifiable connection between the defendant’s actions and the infringement or the likelihood of infringement.”).

Here, however, because the CMI was allegedly removed from an *internal* dataset, that alleged omission has no effect on the public at all. Because the dataset is not publicly accessible, the inclusion of CMI in that dataset could not “inform the public” about the work at issue. *Fischer*, 286 F. Supp. 3d at 610-11. Nor could the alleged exclusion of CMI from that dataset

withhold any such information from public view. Therefore, even if it were an “infringement” for OpenAI to create copies of works of journalism for its training dataset, the removal of CMI from those copies could not plausibly “induce, enable, facilitate, or conceal” that infringement—and OpenAI certainly could not have “know[n]” that it would. 17 U.S.C. § 1202(b). And, in any event, the fact that there are “instances in which author or title information is included in a response” from ChatGPT (Compl. ¶ 40) forecloses any argument that the alleged “removal” of CMI from the training dataset was intended to facilitate or conceal the creation of supposedly infringing ChatGPT outputs. Plaintiff has also failed to plead any facts showing that OpenAI’s alleged removal of CMI was “intentional, rather than merely an unintended side effect” of the training process, for which no Section 1202(b) claim can lie. *See Kelly*, 77 F. Supp. 2d at 1122.

Further, Plaintiff’s only examples of the infringement allegedly induced, enabled, facilitated, or concealed by the removal of CMI from training data consist of alleged “ChatGPT responses that incorporated material from Plaintiff’s copyright-protected works,” “regurgitated copyright-protected works” or “utilized Plaintiff’s copyright-protected works.” (Compl. ¶¶ 57-61.) But Plaintiff never offers an example of a ChatGPT output that “incorporated material from,” “regurgitated,” or “incorporated” any of *Plaintiff’s* unidentified works of journalism, as discussed in Section IV.A, *supra*. In fact, Plaintiff never alleges any specific examples of ChatGPT output *at all*. And if the complaint does not plausibly allege infringing outputs in the first place, it cannot possibly raise a plausible inference that OpenAI should have known about such outputs. *See Corelogic*, 899 F.3d at 674 (rejecting DMCA claim because Plaintiff did not allege that defendant was aware of a “pattern of conduct” or “established modus operandi” giving rise to an inference that defendant was “aware of the probable future impact of its actions” regarding CMI).

5. Plaintiff's Section 1202(b)(3) Claim Fails For Multiple Reasons.

Plaintiff also fails to plausibly allege that OpenAI “distribute[d]” “works” or “copies of works” without CMI and with reason to know that such distribution would induce, enable, facilitate or conceal copyright infringement, for at least three reasons. *See* 17 U.S.C. § 1202(b)(3).

First, Plaintiff does not plausibly allege that a distribution in fact occurred. The complaint recites only that OpenAI “shared copies of Plaintiff’s works” with Microsoft “in connection with the development of ChatGPT.” (Compl. ¶ 65.) To support this contention, Plaintiff gestures towards unspecified “publicly available information regarding the relationship between Defendant Microsoft and the OpenAI Defendants.” (*Id.* ¶ 46.) But Plaintiff does not plausibly explain when, why, or how OpenAI “shared” any of its works of journalism with Microsoft (or even what those works are). Plaintiff does not even identify any of the “publicly available information” on which its allegations depend.

Second, Plaintiff does not allege that the training data OpenAI supposedly shared with Microsoft included identical copies of Plaintiff’s works. Courts have cabined applications of Section 1202 to circumstances in which the works in question were “substantially or entirely reproduced.” *Fischer*, 286 F. Supp. 3d at 609. Accordingly, if OpenAI “shared” anything less than an identical copy of all (or almost all) of one of Plaintiff’s works, that “shar[ing]” would not qualify as a “distribut[ion]” under Section 1202(b)(3). 17 U.S.C. § 1202(b)(3). For example, Plaintiff alleges that its works of journalism were included in a “collection[] of links posted on the website Reddit” or in a “scrape of most of the internet,” which OpenAI then “adapted” into its “ChatGPT training sets.” (Compl. ¶¶ 30-31.) But an allegation that OpenAI “adapted” its data is flatly inconsistent with any contention that the “training sets . . . provided to Microsoft” were “substantial[] or entire[] reproduc[tions]” of Plaintiff’s unidentified works. *Fischer*, 286 F.

Supp. 3d at 609; *see also Robert L. Stark Enters., Inc. v. Neptune Design Grp., LLC*, No. 16-cv-264, 2017 WL 1345195, at *11 (N.D. Ohio Apr. 12, 2017) (rejecting Section 1202(b)(3) claim based on contention that alleged infringer “created derivative works without retaining [opponent’s] CMI” because the statute requires that “copies” be distributed); *Frost-Tsuji Architects v. Highway Inn, Inc.*, No. 13-cv-00496, 2015 WL 263556 at *4 (D. Haw. Jan. 21, 2015) (rejecting Section 1202(b) claim because “[a]t most,” defendant “created [a] derivative work,” which does not support claim that it “removed the [CMI] from [plaintiff’s] original work”). Regardless of whether these facts plausibly state a claim for copyright infringement, they do not state a claim for violation of the DMCA because Plaintiff fails to allege that OpenAI actually shared “works” or “copies of works” under Section 1202(b)(3).

Third, even if OpenAI “shared” identical copies of Plaintiff’s works—which the complaint does not allege—Plaintiff still fails to allege that OpenAI did so with reason to know that it would “induce, enable, facilitate, or conceal” copyright infringement. 17 U.S.C. § 1202(b)(3). As with Plaintiff’s deficient Section 1202(b)(1) claim, this is because Plaintiff does not identify a single instance of infringement of one of its own works, let alone an infringement relating to “shared” training data with Microsoft. Accordingly, Plaintiff has not plausibly alleged facts giving rise to an inference that OpenAI was “aware of the probable future impact of its actions.” *See Corelogic*, 899 F.3d at 675. Additionally, as with the 1202(b)(1) claim, if the “point of CMI” is to provide information to “the public,” *Roberts v. BroadwayHD LLC*, 518 F. Supp. 3d 719, 737 (S.D.N.Y. 2021), it is far from obvious how sharing information *with Microsoft* without CMI could facilitate or conceal infringement.

In sum, because Plaintiff does not plausibly allege a “distribut[ion],” never alleges that OpenAI supposedly “shared” identical copies with Microsoft, and fails to explain how OpenAI

could have known that its alleged “shar[ing]” would induce, enable, facilitate, or conceal copyright infringement, Plaintiff fails to state a claim under Section 1202(b)(3).

V. CONCLUSION

For all of these reasons, Plaintiffs’ complaint fails to state any claim against OpenAI. The complaint should be dismissed in its entirety.

Dated: April 15, 2024

Respectfully submitted,

MORRISON & FOERSTER LLP

By: /s/ Joseph C. Gratz

Joseph C. Gratz (*pro hac vice*)

JGratz@mofocom

Vera Ranieri (*pro hac vice*)

VRanieri@mofocom

425 Market Street

San Francisco, CA 94105-2482

Telephone: 415.268.7000

Facsimile: 415.268.7522

Allyson R. Bennett (*pro hac vice*)

abennett@mofocom

Rose S. Lee (*pro hac vice*)

roselee@mofocom

707 Wilshire Boulevard, Suite 6000

Los Angeles, CA 90017-3543

Telephone: 213.892.5200

Facsimile: 213.892.5454

Attorneys for Defendants

OPENAI, INC., OPENAI GP, LLC,

OPENAI, LLC, OPENAI OP CO LLC,

OPENAI GLOBAL LLC, OAI

CORPORATION, LLC, and OPENAI

HOLDINGS, LLC

LATHAM & WATKINS LLP

By: /s/ Allison L. Stillman

Joseph R. Wetzel
Joseph.wetzel@lw.com
Andrew M. Gass (*pro hac vice*)
andrew.gass@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Allison L. Stillman
alli.stillman@lw.com
Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.751.4864
Attorneys for Defendants
OPENAI, INC., OPENAI GP, LLC,
OPENAI, LLC, OPENAI OPKO LLC,
OPENAI GLOBAL LLC, OAI
CORPORATION, LLC, and OPENAI
HOLDINGS, LLC