

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----	X	
RAW STORY MEDIA, INC., ALTERNET	:	
MEDIA, INC.,	:	NO. 1:24-cv-01514-CM
	:	
Plaintiffs,	:	
	:	
v.	:	
	:	
OPENAI, INC., OPENAI GP, LLC,	:	
OPENAI, LLC, OPENAI OPKO LLC,	:	
OPENAI GLOBAL LLC, OAI	:	
CORPORATION, LLC, and OPENAI	:	
HOLDINGS, LLC,	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

I.	INTRODUCTION	1
II.	PLAINTIFFS’ 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. Plaintiffs Lack Article III Standing.....	4
	1. The mere (alleged) absence of CMI in non-disseminated training data is not a concrete injury-in-fact.	5
	2. Plaintiffs have not otherwise alleged any harm.	7
	B. Plaintiffs Fail To State a Claim Under Section 1202(b) of the DMCA.....	9
	1. Copyright management information and the DMCA	9
	2. Plaintiffs are not among those authorized to sue under 17 U.S.C. § 1203(a).	11
	3. Plaintiffs fail to specify the works at issue.	12
	4. Plaintiffs’ claim fails because they did not adequately plead scienter.	14
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)	12, 16
<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).....	10, 16
<i>Bell Atl. Corp. v. Twombly</i> , 550 U.S. 544 (2007).....	4
<i>Brown v. Showtime Networks, Inc.</i> , 394 F. Supp. 3d 418 (S.D.N.Y. 2019).....	4
<i>Cable v. Agence Fr. Presse</i> , 728 F. Supp. 2d 977 (N.D. Ill. 2010)	10
<i>Carter v. HealthPort Techs., LLC</i> , 822 F.3d 47 (2d Cir. 2016).....	4, 9
<i>Cole v. John Wiley & Sons, Inc.</i> , No. 11-cv-2090, 2012 WL 3133520 (S.D.N.Y. Aug. 1, 2012).....	13
<i>Cortlandt St. Recovery Corp. v. Hellas Telecomms., S.a.r.l.</i> , 790 F.3d 411 (2d Cir. 2015).....	4
<i>Davis v. FEC</i> , 554 U.S. 724 (2008).....	5
<i>Doe 1 v. GitHub, Inc.</i> , 672 F. Supp. 3d 837 (N.D. Cal. 2023)	8, 9
<i>Dow Jones & Co., Inc. v. Int’l Sec. Exch., Inc.</i> , 451 F.3d 295 (2d Cir. 2006).....	13, 14
<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).....	13
<i>First Nationwide Bank v. Gelt Funding Corp.</i> , 27 F.3d 763 (2d Cir. 1994).....	15

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

Free Speech Sys., LLC v. Menzel,
390 F. Supp. 3d 1162 (N.D. Cal. 2019)14

Joint Stock Co. Channel One Russ. Worldwide v. Infomir LLC,
No. 16-cv-1318, 2017 WL 696126 (S.D.N.Y. Feb. 15, 2017)13

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)11, 17

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

Lujan v. Defs. of Wildlife,
504 U.S. 555 (1992).....5, 8

Mango v. BuzzFeed, Inc.,
970 F.3d 167 (2d Cir. 2020).....15

McGucken v. Shutterstock, Inc.,
No. 22-cv-00905, 2023 WL 6390530 (S.D.N.Y. Oct. 2, 2023).....15

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

Ramirez v. TransUnion LLC,
951 F.3d 1008 (9th Cir. 2020)7

Reiffer v. NYC Luxury Limousine Ltd.,
No. 22-cv-2374, 2023 WL 4029400 (S.D.N.Y. June 15, 2023)11

Sherwood 48 Assocs. v. Sony Corp. of Am.,
76 F. App'x 389 (2d Cir. 2003)12

Sitnet LLC v. Meta Platforms, Inc.,
No. 23-cv-6389, 2023 WL 6938283 (S.D.N.Y. Oct. 20, 2023).....12

Steele v. Bongiovi,
784 F. Supp. 2d 94 (D. Mass. 2011)11

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

TransUnion LLC v. Ramirez,
594 U.S. 413 (2021).....2, 5, 6, 7

<i>Tremblay v. OpenAI Inc.</i> , No. 23-cv-03416, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024)	15
<i>Victor Elias Photography, LLC v. Ice Portal, Inc.</i> , 43 F.4th 1313 (11th Cir. 2022)	16
<i>Wolo Mfg. Corp. v. ABC Corp.</i> , 349 F. Supp. 3d 176 (E.D.N.Y. 2018)	13
<i>Zuma Press, Inc. v. Getty Images (US), Inc.</i> , 845 F. App'x. 54 (2d Cir. 2021)	11
Statutes	
17 U.S.C. § 508.....	14
17 U.S.C. § 1202.....	<i>passim</i>
17 U.S.C. § 1203.....	11, 12
Other Authorities	
Cyberspace, 28 CONN. L. REV. 981, 990 (1996).....	10
Fed. R. Civ. P. 8.....	4
Fed. R. Civ. P. 12.....	3, 4
<i>Intellectual Property and the National Information Infrastructure: The Report of the Working Group on Intellectual Property Rights 235–36 (1995)</i>	10
U.S. Constitution Article III.....	<i>passim</i>

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 Plaintiffs want 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 Plaintiffs’ Complaint is that OpenAI allegedly removed author, title, and terms of use information from copies of Plaintiffs’ unspecified “works of journalism” that OpenAI allegedly

used to “train” its large language models. (Compl. (ECF No. 1) ¶ 49.) To support its claim, Plaintiffs aver generally that ChatGPT has “regurgitate[d]” or “mimic[ked]” copyrighted journalism in response to user prompts. (*Id.* ¶¶ 34-35.) But Plaintiffs never offer 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 Plaintiffs’. Indeed, the Complaint fails even to identify Plaintiffs’ “works of journalism” at all. Absent any allegation of concrete harm to Plaintiffs, Raw Story and Alternet Media lack Article III standing under *TransUnion LLC v. Ramirez*, 594 U.S. 413 (2021). And the Complaint otherwise fails to state a claim under the Digital Millenium Copyright Act.

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

II. PLAINTIFFS’ ALLEGATIONS

Plaintiffs are news organizations that publish “copyrighted works of journalism” “on the internet.” (Compl. ¶¶ 7, 31.) They bring one claim against seven OpenAI entities (which they refer to collectively as “the OpenAI defendants,” without differentiating among them) for violation of Section 1202(b)(1) of the Digital Millenium Copyright Act. As relevant here, that section prohibits the “intentional[] remov[al] or alter[ation]” of copyright management information (“CMI”) with “reasonable grounds to know” that the removal “will induce, enable, facilitate, or conceal” copyright infringement. *See* 17 U.S.C. § 1202(b)(1). Plaintiffs do not bring a claim for copyright infringement itself, nor do they provide an example of any of their articles that they believe an OpenAI entity has reproduced without its permission.

Instead, the core of Plaintiffs’ 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. ¶¶ 7, 37-42.) To support these claims, Plaintiffs also contend 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.* ¶¶ 34-35.) But Plaintiffs do not specifically identify a single “work of journalism” that they own—let alone one that was supposedly used by OpenAI for training or contained in a ChatGPT output. Instead, they just say that the works at issue were “published on the internet.” (*Id.* ¶ 31.)

Plaintiffs also allege that “[e]arlier versions of ChatGPT (prior to GPT-4) were trained” on three datasets known as WebText, WebText2, and Common Crawl. (*Id.* ¶ 29.) Plaintiffs allege 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.* ¶ 30.) Plaintiffs, however, do not allege whether any of their works actually appeared in these datasets. Instead, Plaintiffs allege only that “[t]housands” of unidentified examples of its works appeared in “recreated approximations” of the datasets. (*Id.* ¶ 37.) But Plaintiffs offer no further information about who created these “approximations,” how they were created, or which of Plaintiffs’ 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, a plaintiff “must do more” than “merely attach labels and conclusions to bald factual assertions.” *Brown v. Showtime Networks, Inc.*, 394 F. Supp. 3d 418, 436 (S.D.N.Y. 2019) (cleaned up).

IV. ARGUMENT

A. Plaintiffs Lack Article III Standing.

In order to bring a claim, Plaintiffs must plead a cognizable injury to themselves, not to someone else. Plaintiffs allege that ChatGPT has provided responses to users containing material from someone else’s “works of journalism.” (Compl. ¶¶ 34-36.) But Plaintiffs have not pointed to a single example of a ChatGPT output that looks anything like one of their *own* copyrighted works. And the allegation that Plaintiffs’ works merely appear in the training data without CMI, in the absence of an allegation that ChatGPT has actually outputted their works, is not a sufficiently concrete harm. Plaintiffs therefore have failed to plead that they 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, the Supreme Court has explained that Congress cannot itself create Article III standing. That is, even if Congress creates a cause of action by statute, Article III still requires a plaintiff to allege a “concrete” injury by demonstrating a “physical, monetary, or cognizable intangible harm traditionally recognized as providing a basis for a lawsuit in American courts.” *TransUnion LLC v. Ramirez*, 594 U.S. 413, 427 (2021). While Congress may “elevate harms that exist in the real world . . . to actionable legal status, it may not simply enact an injury into existence, using its lawmaking power to transform something that is not remotely harmful into something that is.” *Id.* at 426 (cleaned up). Rather, a plaintiff must identify “a close historical or common-law analogue for their asserted injury” to satisfy Article III’s requirements. *Id.* at 424.

1. The mere (alleged) absence of CMI in non-disseminated training data is not a concrete injury-in-fact.

Plaintiffs allege that OpenAI removed copyright management information (“CMI”) from copies of Plaintiffs’ works of journalism that were “included . . . in training sets used to train ChatGPT” in violation of Section 1202(b)(1) of the DMCA. (Compl. ¶¶ 49–51.) But the alleged presence of these copies in OpenAI’s training data, standing alone, cannot satisfy Article III’s demand to demonstrate a concrete and particularized injury-in-fact. Indeed, the

Supreme Court rejected a strikingly similar allegation of harm in *TransUnion*. That case involved claims under the Fair Credit Reporting Act—another relatively recent federal statute that created a cause of action to sue and recover damages for certain statutory violations. *TransUnion*, 594 U.S. at 418–19. Plaintiffs sued TransUnion—a credit reporting agency—for failing to follow reasonable procedures to ensure the accuracy of credit files that included alerts falsely labeling the plaintiffs as potential terrorists. *Id.* at 430. At the Supreme Court, TransUnion did not dispute that this statutory violation had occurred for each plaintiff. *See id.* at 421.

The Court held, however, that not every plaintiff had standing. Rather, only the plaintiffs whose false credit information was *actually disseminated* to potential creditors had suffered a concrete harm sufficient to confer Article III standing. *Id.* at 432. To be sure, as to the other plaintiffs, those statutory violations “existe[d] . . . in a consumer’s internal credit file at TransUnion.” *Id.* at 433. But the Court explained that the “mere presence of an inaccuracy in an internal credit file, if it is not disclosed to a third party, causes no concrete harm” sufficient for standing. *Id.* at 434.

The Court also held that the risk of future harm, standing alone, “cannot qualify as a concrete harm—at least unless the exposure to the risk of future harm itself causes a *separate* concrete harm.” *Id.* at 436 (emphasis in original). And while a “sufficiently imminent and substantial” risk of future harm may support standing for injunctive relief in some circumstances, the Court explained that the plaintiffs had failed to “demonstrate a sufficient likelihood that their individual credit information would be requested by third-party businesses and provided by TransUnion.” *Id.* at 436; 438. Accordingly, only those plaintiffs whose false credit information was actually provided to third parties had Article III standing to sue under the statute. *Id.* at 433.

Here, Plaintiffs similarly bring a claim based on the alleged presence of data in a nonpublic dataset, in a condition that they claim violates a federal statute. It was not disputed on appeal that the *TransUnion* plaintiffs' credit information sat in TransUnion's internal credit files with inaccuracies, *see Ramirez v. TransUnion LLC*, 951 F.3d 1008, 1017 (9th Cir. 2020); here, Plaintiffs allege that their articles sit in OpenAI's internal training data with missing CMI. But unless Plaintiffs plausibly allege that these copies were somehow *actually shared* with a third party, that injury is not sufficiently concrete for Article III standing. As the Court explained in *TransUnion*, "where allegedly inaccurate or misleading information sits in a company database, the plaintiffs' harm is roughly the same, legally speaking, as if someone wrote a defamatory letter and then stored it in her desk drawer." *Id.* at 434. Accordingly, just as a "letter that is not sent does not harm anyone," neither does the allegedly missing CMI in an internal database harm anyone without an allegation that ChatGPT has outputted Plaintiffs' works to users in the real world—an allegation Plaintiffs have not made, and cannot make. *Id.*

2. Plaintiffs have not otherwise alleged any harm.

Since the absence of CMI in the training data, without more, is insufficient to confer standing, Plaintiffs must otherwise allege some sort of concrete injury to establish this Court's subject matter jurisdiction. But the Complaint contains no other allegation or explanation of whether and how Plaintiffs were harmed. Plaintiffs allege that their journalism is "the result of significant investments by Plaintiffs in the human and other resources necessary to report on the news." (Compl. ¶ 32.) Even so, Plaintiffs have not plausibly alleged that any ChatGPT output has ever resembled one of Plaintiffs' articles. Nor have Plaintiffs offered any plausible allegation of harm to their investments from the mere presence of some unidentified subset of articles as a minuscule fraction of the "scrape of most of the internet" used to train models used for ChatGPT, in the absence of such an output. (*Id.* ¶ 29.)

Instead, the Complaint makes only generalized allegations that ChatGPT has outputted *someone's* copyrighted works of journalism to users. (*See, e.g.*, Compl. ¶¶ 34-35.) But Plaintiffs' allegations do not plausibly suggest that *Plaintiffs'* articles (as opposed to some other new organization's articles) were actually outputted by ChatGPT. Plaintiffs *do not identify a single example* of ChatGPT outputting their "works of journalism," or even anything substantially similar to or derived from their "works of journalism." In fact, the Complaint fails to identify with any reasonable specificity what their "works of journalism" *are*. It states only that those works are "published on the internet." (*Id.* ¶ 31.) And without information about what, exactly, Plaintiffs' 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[s] in a personal and individual way." *Lujan*, 504 U.S. at 560 n.1. Accordingly, any injury connected to ChatGPT outputs is insufficiently particularized to confer Article III standing to Plaintiffs.

Doe 1 v. GitHub, Inc., another DMCA case brought against OpenAI, is instructive on this point. 672 F. Supp. 3d 837 (N.D. Cal. 2023). 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 I*, Plaintiffs do “not allege that they themselves have suffered the injury they describe.” *Id.* at 850. Just as the plaintiffs in *Doe I* anchored their claim on an allegation that AI output has “matched licensed code written by [other] Github user[s],” *id.*, here too Plaintiffs rely 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. ¶¶ 34-35.) The Complaint fails to allege any examples of “regurgitat[ion]” or “mimic[ry]” happening to *Plaintiffs*, nor any reason why it would happen to Plaintiffs. Accordingly, the Complaint fails to allege “facts that affirmatively and plausibly suggest that the plaintiff[s] ha[ve] standing to sue,” *Carter*, 822 F.3d at 56 (cleaned up), and Plaintiffs’ claim should be dismissed.

B. Plaintiffs Fail To State a Claim Under Section 1202(b) of the DMCA.

Plaintiffs bring one claim against OpenAI under Section 1202(b)(1) of the DMCA. But this claim is far afield from the statutory language and purpose of Section 1202, which is briefly surveyed below. That background illuminates the flaws in this claim. As addressed below, Plaintiffs’ claim fails because Plaintiffs (1) are not among the those authorized to sue under the statute, (2) have not identified the works from which OpenAI allegedly removed CMI, and (3) have failed to adequately plead that OpenAI had reason to know its purported CMI removal 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. Section 1202 plaintiffs must also plead and prove both (1) that the removal or alteration was done “intentionally” and (2) that act was performed with “reasonable grounds to know . . . that it 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 Plaintiffs’ DMCA claim.

2. Plaintiffs are not among those authorized to sue under 17 U.S.C. § 1203(a).

Even assuming Plaintiffs had standing under Article III—they do not—Plaintiffs’ Section 1202(b) claim separately fails because Plaintiffs are not within the class of plaintiffs that Congress authorized to sue under that section. “[T]o have standing” to sue for a DMCA violation, Plaintiffs “must show that [they] w[ere] 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, Plaintiffs do not explain how they were injured by OpenAI’s alleged removal of CMI from its training dataset. Even assuming Plaintiffs’ works were included in OpenAI’s training dataset—which Plaintiffs have not plausibly alleged, *see* §IV.B.4., *infra*—Plaintiffs still have 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, they also have not plausibly alleged). This “failure to allege an injury” connected to the alleged DMCA violation is separately “fatal” to Plaintiffs’ claim. *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. Plaintiffs fail to specify the works at issue.

Plaintiffs’ Section 1202 claim is 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 Plaintiffs do 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 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 plaintiffs’ 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, Plaintiffs make no attempt whatsoever to identify the works on which it predicates its Section 1202(b) claim. Plaintiffs make general allegations about its “works of journalism” that were “published on the internet.” (Compl. ¶ 31.) But this is too imprecise to “give [OpenAI] fair notice of the claims against [it].” *See Dow Jones*, 451 F.3d at 307. Indeed, Plaintiffs’ other allegations serve only to highlight this imprecision—according to the Complaint, “thousands of Plaintiffs’ copyrighted works were included in Defendants’ training sets,” but Plaintiffs do not (or cannot) identify a single one of them. (*See* Compl. ¶ 41.)

Even when Plaintiffs filed their “Report on the Filing or Determination of an Action or Appeal Regarding a Copyright” on this Court’s docket, they still failed to identify the works at issue. (*See* ECF No. 7.) 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 their form, Plaintiffs only list “Various” in the fields for “Title of Work” and “Author of Work.” (*Id.*) As a result, not only does OpenAI lack fair notice of the claim 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, Plaintiffs fail to state a Section 1202(b) claim.

4. Plaintiffs’ claim fails because they did not adequately plead scienter.

Plaintiffs also have not pled facts sufficient to meet the DMCA’s “double-scienter” requirement. To plead a violation of Section 1202(b)(1), Plaintiffs 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, Plaintiffs 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)). Plaintiffs’ allegations, however, miss the mark.

As an initial matter, Plaintiffs fail to plausibly allege the intentional removal CMI from any of their works. Plaintiffs allege 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. ¶¶ 34-35.) From that, Plaintiffs speculate that “ChatGPT would have learned to communicate [CMI] when providing responses to users unless Defendants trained it otherwise.” (*Id.* ¶ 38.) 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 Plaintiffs’ unidentified works of journalism actually appeared in the “regurgitate[d]” or “mimic[ked]” outputs (Compl. ¶¶ 34-35), 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, Plaintiffs do 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 Plaintiffs admit that OpenAI’s training dataset is not publicly accessible, (Compl. ¶ 28), 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. ¶ 39) 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. Plaintiffs have 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, Plaintiffs’ 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 Plaintiffs’ copyright-protected works,” “regurgitated copyright-protected works” or “utilized Plaintiffs’ copyright-protected works.” (Compl. ¶¶ 52-56.) But Plaintiffs never offer an example of a ChatGPT output that “incorporated material from,” “regurgitated,” or “incorporated” any of *Plaintiffs’* unidentified works of journalism, as discussed in Section IV.A, *supra*. In fact, Plaintiffs never allege 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).

V. CONCLUSION

For all of these reasons, Plaintiffs' Complaint should be dismissed for lack of Article III standing and failure to state any claim against OpenAI. The Complaint should be dismissed in its entirety.

Dated: April 29, 2024

Respectfully submitted,

MORRISON & FOERSTER LLP

By: /s/ Allyson R. Bennett

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 OPCO LLC,

OPENAI GLOBAL LLC, OAI

CORPORATION, LLC, and OPENAI

HOLDINGS, LLC

LATHAM & WATKINS LLP

By: /s/ Joseph R. Wetzel

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

KEKER, VAN NEST & PETERS LLP

By: /s/ R. James Slaughter

Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com
R. James Slaughter (*pro hac vice*)
rslaughter@keker.com
Paven Malhotra
pmalhotra@keker.com
Michelle Ybarra (*pro hac vice*)
mybarra@keker.com

Nicholas S. Goldberg (*pro hac vice*)
ngoldberg@keker.com
Thomas E. Gorman (*pro hac vice*)
tgorman@keker.com
Katie Lynn Joyce (*pro hac vice*)
kjoyce@keker.com
633 Battery Street
San Francisco, CA 94111-1809
Telephone: 415.391.5400
Facsimile: 415.397.7188

Attorneys for Defendants
OPENAI, INC., OPENAI GP, LLC,
OPENAI, LLC, OPENAI OPKO LLC,
OPENAI GLOBAL LLC, OAI
CORPORATION, LLC, and OPENAI
HOLDINGS, LLC