

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

RAW STORY MEDIA, INC., ALTERNET
MEDIA, INC.,

Plaintiffs,

v.

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

Defendants.

No. 1:24-cv-01514-CM

**PLAINTIFFS' MEMORANDUM OF LAW IN SUPPORT OF
THEIR MOTION FOR LEAVE TO AMEND COMPLAINT OR, IN THE
ALTERNATIVE, TO CONTINUE TAKING JURISDICTIONAL DISCOVERY**

TABLE OF CONTENTS

I. INTRODUCTION.....1

II. BACKGROUND2

III. ARGUMENT.....3

A. The Court should grant Plaintiffs leave to amend their complaint.....3

1. Leave to amend is freely given.....3

2. The proposed amendments present a close analogy between copyright infringement and Plaintiffs’ DMCA injuries based on the specific facts of this case3

3. The proposed amendments also justify standing based on disgorgement of the unlawful profits Defendants obtained by removing Plaintiffs’ CMI.....6

B. In the alternative, the Court should allow Plaintiffs to continue their jurisdictional discovery into Defendants’ dissemination of their articles.....10

C. The proposed amendments resolve Defendants’ arguments for dismissal under Rule 12(b)(6).....12

1. Plaintiffs have statutory standing.....12

2. Plaintiffs identify the works at issue.....12

3. Plaintiffs allege that Defendants intentionally removed CMI13

4. Plaintiffs allege that Defendants removed CMI with knowledge, or reasonable grounds to know, that it would induce, enable, facilitate, or conceal an infringement14

IV. CONCLUSION18

TABLE OF AUTHORITIES

	<i>Page</i>
 Cases	
<i>Aaberg v. Francesca’s Collections, Inc.</i> , No. 17-cv-115, 2018 WL 1583037 (S.D.N.Y. Mar. 27, 2018)	14
<i>Amidax Trading Grp. v. S.W.I.F.T. SCRL</i> , 671 F.3d 140 (2d Cir. 2011)	10
<i>Atrium Grp. De Ediciones Y Publicaciones, S.L. v. Harry N. Abrams, Inc.</i> , 565 F. Supp. 2d 505 (S.D.N.Y. 2008).....	8
<i>Barney v. Saunders</i> , 57 U.S. (16 How.) 535, (1853)	6
<i>Devocean Jewelry LLC v. Associated Newspapers Ltd.</i> , No. 16-cv-2150, 2016 WL 6135662 (S.D.N.Y. Oct. 19, 2016).....	15
<i>Dowling v. U.S.</i> , 473 U.S. 207 (1985)	5
<i>F. W. Woolworth Co. v. Contemp. Arts</i> , 344 U.S. 228 (1952).....	5
<i>Fair Isaac Corp. v. Fed. Ins. Co.</i> , 468 F. Supp. 3d 1110 (D. Minn. 2020)	8
<i>Gary Friedrich Enters., LLC v. Marvel Enters., Inc.</i> , 713 F. Supp. 2d 215 (S.D.N.Y. 2010).....	5
<i>Glen v. Am. Airlines, Inc.</i> , 7 F.4th 331 (5th Cir. 2021).....	7
<i>Hirsch v. CBS Broad. Inc.</i> , No. 17-cv-1860, 2017 WL 3393845 (S.D.N.Y. Aug. 4, 2017).....	14
<i>Jewell-La Salle Realty Co. v. Buck</i> , 283 U.S. 202 (1931)	5
<i>Kamen v. Am. Tel. & Tel. Co.</i> , 791 F.2d 1006 (2d Cir. 1986)	10
<i>Marsh & McLennan Agency LLC v. Williams</i> , No. 22-cv-8920, 2023 WL 4534984 (S.D.N.Y. July 13, 2023)	10
<i>Mercer v. Jericho Hotels, LLC</i> , No. 19-cv-5604, 2019 WL 6117317 (S.D.N.Y. Nov. 18, 2019).....	10
<i>Meyer v. Seidel</i> , 89 F.4th 117 (2d Cir. 2023).....	3
<i>Mills 2011 LLC v. Synovus Bank</i> , 921 F. Supp. 2d 219 (S.D.N.Y. 2013)	11
<i>Murphy v. Millennium Radio Grp. LLC</i> , No. 08-cv-1743, 2015 WL 419884 (D.N.J. Jan. 30, 2015)	15

Nintendo of Am., Inc. v. Dragon Pac. Int’l, 40 F.3d 1007 (9th Cir. 1994).....8

Packer on behalf of 1-800-Flowers.Com, Inc. v. Raging Cap. Mgmt., LLC, 105 F.4th 46 (2d Cir. 2024).....6

Red Apple Media, Inc. v. Batchelor, No. 23-cv-10253, 2024 WL 1504439 (S.D.N.Y. Apr. 8, 2024).....8

Saba Capital Cef Opportunities I, Ltd. v. Nuveen Floating Rate Income Fund, 88 F.4th 103 (2d Cir. 2023)5

Sammons v. Colonial Press, 126 F.2d 341 (1st Cir. 1942).....8

TransUnion LLC v. Ramirez, 594 U.S. 413 (2021)4, 6, 8

Tutor Perini Bldg. Corp. v. New York City Reg’l Ctr., LLC, No. 20-cv-731, 2020 WL 7711629 (S.D.N.Y. Dec. 29, 2020)10

Vermont Agency of Nat. Res. v. U.S. ex rel. Stevens, 529 U.S. 765 (2000).....8

Statutes

17 U.S.C. § 1202(c)(2).....6

Other Authorities

Act of May 31, 1790, ch. 15, § 25

Fed. R. Civ. P. 15(a)(2).....3

Restatement (Third) of Restitution and Unjust Enrichment7

S. Rep. 105-190 (1998).....12

I. INTRODUCTION

Defendants injured Plaintiffs by removing CMI from tens of thousands of Plaintiffs' news articles. Doing so enabled Defendants to more efficiently "train" their algorithms on Plaintiffs' work product, creating a company valued at more than \$100 billion. In removing the CMI, Defendants inflicted that injury by performing acts that are not just arguably *analogous* to copyright infringement in the abstract—an analogy the Court rejected—but that *are* copyright infringement under the specific facts of this case. And in removing Plaintiffs' CMI, Defendants also unjustly enriched themselves by shortcutting the need for certain kinds of costly LLM training. Either injury—copyright infringement or unjust enrichment—suffices for Article III standing, and neither was alleged in Plaintiffs' initial Complaint. The proposed amendments allege both and thus are not futile.

If the Court disagrees, and remains convinced that Article III requires dissemination in the context of this case, Plaintiffs respectfully request leave to continue the discovery they have been seeking since April into ChatGPT's dissemination of their articles. Knowledge of dissemination is uniquely within Defendants' possession, Plaintiffs have several colorable bases for believing Defendants disseminated their articles, and Plaintiffs would have already had the evidence needed to plead the point more thoroughly were it not for Defendants' delays in responding to their document requests.

As discussed below, Plaintiffs' proposed amendments also address all the grounds on which Defendants sought dismissal under Rule 12(b)(6), arguments the Court did not yet reach. Indeed, when the *Intercept* plaintiff amended its complaint to add similar allegations, Defendants declined to press many of them. So the proposed amendments would not be futile for this reason, too.

II. BACKGROUND

On February 28, 2024, Plaintiffs filed this action alleging that Defendants violated Section 1202(b)(1) by removing author, title, and copyright notice information from copies of Plaintiffs' copyright-protected news articles used to train ChatGPT. *See generally* Complaint (Feb. 28, 2024), ECF No. 1. The Court set an initial pretrial conference date for April 18, 2024. *See* Order Scheduling an Initial Pretrial Conference (Mar. 6, 2024), ECF No. 16. Four days after the conference, Plaintiffs served their first set of discovery requests. These requests sought, *inter alia*, all chat responses that either regurgitated or included any portion of Plaintiffs' articles. *See* Exhibit A (Proposed First Amended Complaint, or "PFAC") ¶ 77. Despite Plaintiffs' diligent efforts, Defendants have not responded to these requests. *Id.* ¶ 78.

Defendants moved to dismiss the Complaint under Fed. R. Civ. P. 12(b)(1) and Fed. R. Civ. P. 12(b)(6). *See generally* Memorandum of Law in Support of Defendants' Motion to Dismiss (Apr. 29, 2024), ECF No. 68. In opposing Defendants' Rule 12(b)(1) motion, Plaintiffs argued that there is a sufficiently close analogy between DMCA violations and copyright infringement to satisfy the Article III standards announced in *TransUnion LLC v. Ramirez*, 594 U.S. 413 (2021). To defend that analogy, Plaintiffs highlighted similarities in how Congress structured the DMCA and the Copyright Act and argued that "both the Copyright Act and the DMCA protect similar rights involving copyright-protected works," such as the right to reproduce works and prepare derivatives. Plaintiffs' Memorandum of Law in Opposition to Defendants' Motion to Dismiss, at 6 (May 13, 2024), ECF No. 70 ("MTD Opp."). Plaintiffs also sought leave to replead should the Court identify any deficiencies with their Complaint. *See id.* at 21.

The Court granted Defendants' motion under Rule 12(b)(1). It reasoned, primarily, that copyright infringement differs from injuries under the DMCA in that the former protects owners' rights to decide what may happen with future iterations of the work—such as who may copy it—

while the latter “protects copyright owners from specified interferences with the *integrity of a work’s CMI*.” Decision and Order, at 6 (Nov. 7, 2024), ECF No. 117 (“MTD Order”) (emphasis in original). Thus, the Court held, it is possible for a defendant to create infringing copies without violating Section 1202(b)(1) “as long as Defendants keep Plaintiffs’ CMI intact.” *Id.* The Court thus was not “convinced that the mere removal of identifying information from a copyrighted work—absent dissemination—has *any* historical or common-law analogue.” *Id.* at 7 (emphasis in original). Ultimately, the Court granted Defendants’ motion to dismiss. It also denied Plaintiffs’ request for leave to replead without prejudice to renewal on a proper record, leaving open that “another statute or legal theory” could ground standing, but that “that question is not before the Court today.” *Id.* at 9; *see also id.* at 10.

III. ARGUMENT

A. The Court should grant Plaintiffs leave to amend their complaint.

1. Leave to amend is freely given.

Motions for leave to amend are governed by Rule 15, which provides that “[t]he court should freely give leave when justice so requires.” Fed. R. Civ. P. 15(a)(2). As here, “[w]here the possibility exists that the defect can be cured and there is no prejudice to the defendant, leave to amend at least once should normally be granted as a matter of course.” *Meyer v. Seidel*, 89 F.4th 117, 140 (2d Cir. 2023). The Court need not grant leave to amend, however, when the amendment would be futile. *See id.*

2. The proposed amendments present a close analogy between copyright infringement and Plaintiffs’ DMCA injuries based on the specific facts of this case

In rejecting the analogy between injuries under the Copyright Act and the DMCA, the Court reasoned, at least in part, that a defendant can violate one while complying with the other, such as by creating an infringing copy that preserves the original’s CMI. *See* MTD Order at 6.

The Court's holding entails that there is no *per se* analogy between injuries under the two statutes. But it leaves open the possibility of such an analogy in discrete cases, where the violations go hand in hand.

This is such a case. If permitted to replead, Plaintiffs will allege that Defendants removed Plaintiffs' CMI through a technical process that constitutes actual *prima facie* copyright infringement. In particular, Defendants downloaded Plaintiffs' copyright-protected works and applied computer algorithms to create further copies with CMI deliberately removed. *See* PFAC ¶¶ 42-52. Defendants cannot reasonably dispute that such copying is *prima facie* copyright infringement. Indeed, they did not dispute it when other news organizations sued them for infringing their copyright, electing instead to stake their defense on fair use.¹ So even if DMCA violations are not always analogous to copyright infringement, they are on the particular facts of this case, where Defendants committed actual *prima facie* copyright infringement in the course of violating the DMCA. And the Supreme Court made clear in *TransUnion* that what governs is not the elements of the legal claims, but rather the specific injury in fact that has been alleged. Thus, claims for the same statutory violations survived a standing challenge when they included the additional fact that misleading information had been disseminated, even though dissemination was not an element of the statutory violation. *See TransUnion*, 594 U.S. at 418-19 (articulating statutory requirements, which does not include dissemination); *id* at 432-34 (allowing plaintiffs' claims to go forward depending on whether the defendants allegedly published their credit files).

¹ *See, e.g.*, Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss, at 2-3, *The New York Times Company v. Microsoft Corp.*, No. 23-cv-11195 (S.D.N.Y. Feb. 26, 2024), ECF No. 52; Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss, at 2, *Daily News, LP v. Microsoft Corp.*, No. 24-cv-3285 (S.D.N.Y. June 11, 2024), ECF No. 82.

Second Circuit case law fully supports standing for the allegations in the Proposed First Amended Complaint. On the facts of this case, the Copyright Act-DMCA analogy is far closer than that between (1) encumbering voting shares through a fund's amendment to its bylaws and (2) common law trespass to chattels, which the Second Circuit found sufficient in *Saba Capital Cef Opportunities 1, Ltd. v. Nuveen Floating Rate Income Fund*, 88 F.4th 103 (2d Cir. 2023). Unlike with Plaintiffs' Section 1202(b)(1) claim, it is doubtful the injury in *Nuveen* could even **theoretically** satisfy the common law tort. Trespass to chattels requires the defendant to assume "physical control" of the chattel, which does not occur when a fund amends its bylaws. *Gary Friedrich Enters., LLC v. Marvel Enters., Inc.*, 713 F. Supp. 2d 215, 231 (S.D.N.Y. 2010) (quoting *Dowling v. U.S.*, 473 U.S. 207, 217 (1985)). It was instead enough that the bylaw amendment caused something "analogous to a property-based injury." *Nuveen*, 88 F.4th at 116. Here, Plaintiffs' DMCA injuries might well **be** copyright infringement. The analogy to copyright infringement is therefore close enough to satisfy Article III.

The infringing nature of the DMCA violation also makes Plaintiffs' injury "real, and not abstract." *TransUnion*, 594 U.S. at 424. Article III draws that line based on whether the injury resembles one "recognized as providing a basis for a lawsuit in American courts." *Id.* Copyright infringement has always been a considered a real-world injury, including when the infringement has no tangible results apart from the infringement itself. *See F. W. Woolworth Co. v. Contemp. Arts*, 344 U.S. 228, 233 (1952) (holding damages available for acts that, apart from the infringement itself, are "uninjurious and unprofitable"); *Jewell-La Salle Realty Co. v. Buck*, 283 U.S. 202, 208 (1931) (construing Copyright Act to mandate statutory damages even if "there is no showing as to actual loss"); Act of May 31, 1790, ch. 15, § 2 (providing infringement damages of 50 cents per page without any other required showing). And this is not surprising given that Article

III grants standing for precisely these kinds of “intangible harms.” *TransUnion*, 594 U.S. at 425. Thus, because Defendants violated the DMCA through acts that also constitute copyright infringement that has been actionable throughout the history of this country, Article III grants authority to seek redress.

3. The proposed amendments also justify standing based on disgorgement of the unlawful profits Defendants obtained by removing Plaintiffs’ CMI.

The Court should also permit the proposed amendments because they support an additional, independently sufficient standing theory: one based on the unjust enrichment Defendants realized by removing Plaintiffs’ CMI, the specifics of which are discussed below. The DMCA recognizes a distinct harm analogous to unjust enrichment by allowing plaintiffs to disgorge “any profits of the violator that are attributable to the violation,” whether or not accompanied by separate monetary damages. 17 U.S.C. § 1202(c)(2). Those harms are sufficiently analogous for Article III purposes.

Disgorgement of ill-gotten profits, even absent economic loss to a plaintiff, can ground a concrete injury after *TransUnion*. The Second Circuit concluded as much in *Packer on behalf of 1-800-Flowers.Com, Inc. v. Raging Cap. Mgmt., LLC*, 105 F.4th 46, 51-56 (2d Cir. 2024), cert. denied, No. 24-408, 2024 WL 4743106 (Nov. 12, 2024), where it held that a shareholder had standing to bring a derivative suit under Section 16(b) of the Securities Exchange Act of 1934 to recover unlawful short swing profits—profits the court did not identify with any economic loss to the company. The court reasoned that a Section 16(b) violation was analogous to the common-law injury of breach of fiduciary duty, where the fiduciary “must account to the beneficiary ‘for all the gain which he has made.’” *Id.* at 54 (quoting *Barney v. Saunders*, 57 U.S. (16 How.) 535, 543, (1853)). So under Second Circuit precedent, standing exists when a defendant breaches a duty and unlawfully profits from the breach.

Unjust enrichment shares that structure. According to the applicable Restatement, “[a] person who is unjustly enriched at the expense of another is subject to liability.” Restatement (Third) of Restitution and Unjust Enrichment, § 1. Critically, “at the expense of another” does not require an “observable loss,” but can instead involve the “violation of the other’s legally protected rights, without the need to show that the claimant has suffered a loss.” *Id.* § 1 cmt. a. (cleaned up). So unjust enrichment has the same breach-plus-profit structure that grounds standing under *Packer*. See also *Glen v. Am. Airlines, Inc.*, 7 F.4th 331, 334 (5th Cir. 2021) (holding that unjust enrichment “has indisputable common-law roots” under *TransUnion*). Indeed, the breach of fiduciary duty that grounded standing in *Packer* is simply a special case of unjust enrichment, further bolstering the similarity between the two. See Restatement (Third) of Restitution and Unjust Enrichment, § 43 cmt. a (“Gain resulting from breach of fiduciary duty is a prime example of the unjust enrichment that the law of restitution condemns.”).

The DMCA violation alleged here led to unlawful profits that fit comfortably within the rubric of unjust enrichment and is thus analogous to it under Article III. According to the Third Restatement, the “legally protected rights” derive from “other legal rules.” *Id.* § 44 cmt. a. Those rules include federal statutes regulating the “misappropriation or infringement of another person’s legally protected rights in any ... information.” *Id.* § 42; see also *id.* § 42 cmt. a (noting that “[t]he principal rights within the scope of this section are defined and regulated by federal statute”). And the DMCA’s prohibition against removing CMI plainly protect a plaintiff’s “information.”

Further, Congress recognized this analogy by enacting a disgorgement provision in the DMCA that mirrors Section 504(b) of the Copyright Act, which courts have recognized to protect

against unjust enrichment.² See, e.g., *Sammons v. Colonial Press*, 126 F.2d 341, 345 (1st Cir. 1942) (“The theory was that it was unconscionable for an infringer to retain a benefit which he had received by the appropriation and use of the plaintiff’s property right; and to prevent unjust enrichment the infringer was treated as a trustee ex maleficio of his ill gotten gains.”); *Nintendo of Am., Inc. v. Dragon Pac. Int’l*, 40 F.3d 1007, 1011 (9th Cir. 1994) (recognizing that profits under the Copyright Act exist “to prevent the defendant’s unjust enrichment”); *Fair Isaac Corp. v. Fed. Ins. Co.*, 468 F. Supp. 3d 1110, 1116 (D. Minn. 2020) (holding “unjust enrichment” among the purposes of Section 504(b)’s disgorgement provision, and noting that actual damages are not necessary for an award of profits). That judgment, while not dispositive, is afforded “due respect.” *TransUnion*, 594 U.S. at 425.

The DMCA’s profit provision, moreover, is unlike the statute *TransUnion* found objectionable. That statute granted statutory damages to any plaintiff about whom the defendant maintained an inaccurate credit file, whether or not it had any real-world effect. See *TransUnion*, 594 U.S. at 418-19. So the Court was concerned that recognizing standing there would allow Congress to create standing in any case merely by authorizing statutory damages. See *id.* at 428; see also *Vermont Agency of Nat. Res. v. U.S. ex rel. Stevens*, 529 U.S. 765, 773 (2000) (holding statutory damages could not ground standing for violations of the False Claims Act because they are merely a “byproduct of the suit”) (cleaned up). But profits are unlike statutory damages. Unlawful profits *are* the real-world effect and are not, like statutory damages, a “byproduct” of the suit. For this reason, conditioning standing on profits will not open the floodgates. A credit

² Indeed, due to the similarity of the two, courts have routinely held state law unjust enrichment claims preempted by the Copyright Act. See, e.g., *Red Apple Media, Inc. v. Batchelor*, No. 23-cv-10253, 2024 WL 1504439, at *9 (S.D.N.Y. Apr. 8, 2024) (citing Section 504(b)’s profits provision to support that conclusion); *Atrium Grp. De Ediciones Y Publicaciones, S.L. v. Harry N. Abrams, Inc.*, 565 F. Supp. 2d 505, 509 (S.D.N.Y. 2008) (holding unjust enrichment preempted).

reporting agency that does nothing with an inaccurate credit file will turn no profit, and the file's subject will lack standing. So too for DMCA violations. Simply removing CMI does not realize a profit and will not, on this theory, ground Article III standing by itself. The plaintiff must allege more.³

If allowed to replead, Plaintiffs will allege more: that Defendants unjustly profited by removing Plaintiffs' CMI. In particular, removing Plaintiffs' CMI, and using the CMI-less copies to train ChatGPT, allowed Defendants to avoid certain types of model training that likely would have cost substantial sums of money. Defendants want ChatGPT to mimic how ordinary English writers express themselves. PFAC ¶ 95. Had Defendants chosen to include CMI in their training sets, ChatGPT would have learned that English speakers include CMI in situations where they do not. *Id.* For instance, were ChatGPT trained largely on content ending with a copyright notice, it would learn that English writers typically put copyright notices at the end of sentences—which they do not except when ascribing copyright. *Id.* The only alternative would be to retrain ChatGPT not to emit copyright notices, which is more resource-intensive and expensive than removing CMI from the start. *Id.* Thus, by removing CMI, Defendants saved retraining resources—a type of unjust enrichment. *See* Restatement (Third) of Restitution and Unjust Enrichment, § 1 cmt. d (noting unjust enrichment applies in cases of “[a] saved expenditure”). That unlawful profit grounds Plaintiffs' standing and provides a second basis to allow Plaintiffs' claims to go forward.

³ On this theory of standing, the plaintiff must ultimately prove that the defendant profited from removing the CMI. But once the plaintiff has established standing, it is not limited to recovering those profits. It may instead elect statutory damages under 17 U.S.C. § 1203(c)(3)(B). Such a plaintiff is similarly situated to one that establishes standing by showing economic loss but elects statutory damages over the value of that loss.

B. In the alternative, the Court should allow Plaintiffs to continue their jurisdictional discovery into Defendants' dissemination of their articles.

The proposed amendments allege all the facts necessary to support standing under either of the above two theories. If the Court disagrees, and maintains that standing requires dissemination, then it should permit Plaintiffs to complete the jurisdictional discovery they have been taking on that issue.

Discovery is often an appropriate device for assessing subject matter jurisdiction. *See Amidax Trading Grp. v. S.W.I.F.T. SCRL*, 671 F.3d 140, 149 (2d Cir. 2011) (holding that “a court should take care to give the plaintiff ample opportunity to secure and present evidence relevant to the existence of jurisdiction”). It is especially appropriate “where the facts are peculiarly within the knowledge of the opposing party.” *Kamen v. Am. Tel. & Tel. Co.*, 791 F.2d 1006, 1011 (2d Cir. 1986). *Compare, e.g., Amidax*, 671 F.3d at 149 (affirming denial of jurisdictional discovery because the plaintiff, “not the defendants, is in control of the relevant jurisdictional evidence”) with *Mercer v. Jericho Hotels, LLC*, No. 19-cv-5604, 2019 WL 6117317, at *3 (S.D.N.Y. Nov. 18, 2019) (granting jurisdictional discovery on voluntary cessation exception to mootness because only the defendant knew its “substantive plans” for the future). Because standing is an aspect of subject matter jurisdiction, courts may order discovery to assess a plaintiff’s standing. *See Marsh & McLennan Agency LLC v. Williams*, No. 22-cv-8920, 2023 WL 4534984, *1 (S.D.N.Y. July 13, 2023) (ordering jurisdictional discovery on standing because “it appears that the evidence solely in the possession of Defendants may be relevant to Plaintiff’s ability to respond to Defendants’ motion and potentially establish standing”). And the standard for ordering jurisdictional discovery is low. It is warranted “[w]hen a plaintiff has made a ‘threshold showing for some basis to assert jurisdiction, such as facts that would support a colorable claim of jurisdiction,’” but where the complaint’s allegations do not establish jurisdiction. *Tutor Perini Bldg. Corp. v. New York City*

Reg'l Ctr., LLC, No. 20-cv-731, 2020 WL 7711629, at *3 (S.D.N.Y. Dec. 29, 2020) (quoting *Mills 2011 LLC v. Synovus Bank*, 921 F. Supp. 2d 219, 228 (S.D.N.Y. 2013)).

If Plaintiffs' standing requires dissemination, jurisdictional discovery is warranted. The relevant fact to establish dissemination—whether Defendants disseminated Plaintiffs' articles—is uniquely within Defendants' possession. It is not feasible for Plaintiffs to identify which articles were disseminated to which users. But Defendants have a repository of all ChatGPT prompts and responses except when users affirmatively disable that feature. PFAC ¶ 64. At the same time, Plaintiffs have alleged at least a colorable basis for believing their articles—tens of thousands of which are in ChatGPT's training sets—have been disseminated. Defendants have admitted that ChatGPT regurgitates material, Copyleaks has found that ChatGPT plagiarizes with alarming frequency, and other news organizations have shown that ChatGPT can disseminate news content. *Id.* ¶¶ 5, 65-68. While Plaintiffs have not shown such dissemination, that is likely because Defendants have recently modified ChatGPT to reduce them—Plaintiffs' counsel were unable to reproduce the same ChatGPT regurgitations, using the same methods, as the *Daily News* plaintiffs did only a few months ago, which according to ChatGPT was for copyright reasons. *Id.* ¶ 67. It is thus at least “colorable” that ChatGPT disseminated Plaintiffs' articles before making these modifications but within the statute of limitations.

Further, Plaintiffs already served the relevant jurisdictional discovery in April: requests for production of ChatGPT responses that regurgitated or incorporated material from Plaintiffs' articles. *Id.* ¶ 77. Thus, in the context of this case, ordering jurisdictional discovery simply means letting some ongoing discovery play out.⁴ Had Defendants timely produced the documents

⁴ Plaintiffs have no objection to a stay of the rest of the pending discovery while the Court decides the jurisdictional issues.

showing dissemination, Plaintiffs would have been able to incorporate them into their proposed amended complaint. It would be unfair to deny Plaintiffs that opportunity solely based on Defendants' delay.

C. The proposed amendments resolve Defendants' arguments for dismissal under Rule 12(b)(6).

Because it dismissed the Complaint for lack of standing, the Court did not decide Defendants' motion to dismiss for failure to state a claim. While none of Defendants' arguments had merit to begin with, they are even weaker in light of the proposed amendments. Thus, the amendments would not be futile as to Defendants' Rule 12(b)(6) motion. Defendants raised four such arguments.

1. Plaintiffs have statutory standing.

Defendants argued under Rule 12(b)(6) that Plaintiffs lack statutory standing under 17 U.S.C. § 1203(a), which provides, "Any person injured by a violation of section 1201 or 1202 may bring a civil action." MTD at 11-12, ECF No. 68. For the reasons given in Plaintiffs' opening brief, MTD Opp. at 10-11, that argument fails as a legal matter because "injured" under Section 1203(a) simply refers to an Article III injury. This is confirmed by the legislative history, which states that this section "sets forth the general proposition that civil remedies are available for violations of sections 1201 and 1202." S. Rep. 105-190, at 38 (1998). On top of that, the proposed amendments clearly allege the type of injury Defendants say would be required. Defendants injured Plaintiffs by copying their articles so as to remove CMI, and by depriving them of unlawful profits they made through removing that CMI. *See* Section III.A, *supra*.

2. Plaintiffs identify the works at issue.

Defendants next argued that the Complaint should be dismissed because it did not identify "any works from which CMI was allegedly removed." MTD at 12. Plaintiffs have now done that.

They have identified all the works they own, published on their web domains, that Defendants likely scraped when creating their training sets. *See* PFAC ¶¶ 28, 29; PFAC Exs. 1, 2. They also provided more detail about training sets Defendants publicly admitted to using, including one called WebText. *Id.* ¶¶ 36-41. They alleged that professional computer scientists recreated WebText and published a list of every URL their recreation—called OpenWebText—contains. *Id.* ¶ 40. That list includes 40,403 articles from Raw Story and 34,003 articles from Alternet, all of which are referenced in Exhibits 3 and 4 to the Proposed First Amended Complaint. *Id.*

Defendants abandoned their work-identification argument after The Intercept did the same in amending its complaint, which, according to Defendants “asserted virtually identical claims based on almost identical allegations to those at issue in this case.” Defendants’ Notice of Supplemental Authority at 1, (S.D.N.Y. June 10, 2024), ECF No. 78.⁵ But Plaintiffs here have done more. Through discovery, they identified many of their works contained in certain of Defendants training sets, and attached a list of those works as an exhibit to the Proposed First Amended Complaint. *See* PFAC ¶ 59; PFAC Exs. 8-9. All this certainly suffices to identify the works at issue.

3. Plaintiffs allege that Defendants intentionally removed CMI.

Defendants also argued that Plaintiffs failed to allege the first element of the DMCA’s “double-scienter” requirement: “that OpenAI intentionally removed CMI.” MTD at 14. In response, Plaintiffs now allege—based on Defendants’ public admissions—that when assembling

⁵ *See* Memorandum of Law in Support of OpenAI Defendants’ Motion to Dismiss, at 10-12, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-1515 (S.D.N.Y. Apr. 15, 2024), ECF No. 53 (arguing plaintiff failed to specify works at issue); First Amended Complaint, ¶¶ 31-32, 39-44, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-1515 (S.D.N.Y. June 21, 2024), ECF No. 87 (identifying works on plaintiff’s website); Memorandum of Law in Support of OpenAI Defendants’ Renewed Motion to Dismiss First Amended Complaint, at 19-22, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-1515 (S.D.N.Y. July 8, 2024), ECF No. 89 (not arguing for dismissal on this ground).

copies of websites for use in their training sets, Defendants applied computer algorithms, called Dagnet and Newspaper, that are expressly designed not to include CMI from the text they extract from the articles, which is then used to train ChatGPT. PFAC ¶¶ 42-52. To verify that these algorithms removed CMI, a data scientist employed by Plaintiffs' counsel applied them to a sample of Plaintiffs' articles. The results, contained in Exhibits 5 and 6 to the Proposed First Amended Complaint, lack the author, title, and copyright notice information with which the articles were originally conveyed. PFAC ¶¶ 48-49; PFAC Exs. 5-6. Further, OpenAI employs highly skilled data scientists who would have known how the algorithm works, and as such Plaintiffs have adequately alleged intent. PFAC ¶ 47. Defendants dropped this argument, too, after *The Intercept* added similar allegations.⁶

4. Plaintiffs allege that Defendants removed CMI with knowledge, or reasonable grounds to know, that it would induce, enable, facilitate, or conceal an infringement.

If permitted to plead, Plaintiffs will supplement their allegations on the second scienter element: that defendants knew, or had reasonable grounds to know, that it would induce, enable, facilitate, or conceal an infringement. The proposed amendments will articulate five theories of scienter, each sufficient to clear the low bar required for alleging scienter under the DMCA. *Aaberg v. Francesca's Collections, Inc.*, No. 17-cv-115, 2018 WL 1583037, at *9 (S.D.N.Y. Mar. 27, 2018); *Hirsch v. CBS Broad. Inc.*, No. 17-cv-1860, 2017 WL 3393845, at *8 (S.D.N.Y. Aug.

⁶ See Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss, at 13, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-1515 (S.D.N.Y. Apr. 15, 2024), ECF No. 53 (arguing plaintiff failed to specify works at issue); First Amended Complaint, ¶¶ 45-54, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-1515 (S.D.N.Y. June 21, 2024), ECF No. 87 (identifying works on plaintiff's website); Memorandum of Law in Support of OpenAI Defendants' Renewed Motion to Dismiss First Amended Complaint, at 19-22, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-1515 (S.D.N.Y. July 8, 2024), ECF No. 89 (not arguing for dismissal on this ground).

4, 2017); *Devocean Jewelry LLC v. Associated Newspapers Ltd.*, No. 16-cv-2150, 2016 WL 6135662, at *2 (S.D.N.Y. Oct. 19, 2016).

Concealing Defendants’ training-based infringement from their users. Defendants infringed Plaintiffs’ copyright by downloading copies of Plaintiffs’ articles into their training sets and using those copies to train ChatGPT, all without permission. PFAC ¶ 62. Defendants thus had reason to know that removing CMI during training would conceal this training-based infringement from users. A ChatGPT response that ***contained*** Plaintiffs’ CMI would inform the user that the response blossomed out of an infringing copy in the training set. After all ChatGPT’s, “responses are the product of its training sets.” *Id.* ¶ 89. Thus, a response that ***omits*** CMI conceals that training-based infringement. And, critically, ChatGPT omits CMI precisely because Defendants removed the CMI during training: if it was trained on CMI, it would have produced CMI. *Id.* ¶ 81.

It is immaterial that Plaintiffs have not (yet) alleged that ChatGPT disseminated their articles in particular. Defendants admit that ChatGPT regurgitates, which is supported by a study showing that it plagiarizes with alarming frequency—containing text identical to pre-existing content over 45 of the time. *Id.* ¶¶ 5, 67. Thus, Defendants at least had reasonable grounds to know that ChatGPT would produce text that conceals input-based infringement from users. And, of course, the DMCA violation is complete at the moment the defendant removed the CMI, whether or not a later infringement—or, necessarily, a later concealment of that infringement—actually results. *See Murphy v. Millennium Radio Grp. LLC*, No. 08-cv-1743, 2015 WL 419884, at *4-5 (D.N.J. Jan. 30, 2015). ChatGPT is a serial plagiarist and had tens of thousands of Plaintiffs’ articles to plagiarize. So Defendants at least had reason to know that it would plagiarize

Plaintiffs' articles and omit the CMI that would otherwise convey their training-based infringement.

Concealing Defendants' output-based infringement from their users. Plaintiffs will also allege that removing CMI during training concealed the infringement Defendants had reasonable grounds to know that they would commit by producing unauthorized regurgitations and abridgements. Had Defendants included the CMI in their training data, ChatGPT would have produced that CMI in its regurgitations and abridgements. PFAC ¶ 81. This in turn conceals from users that ChatGPT output infringes Plaintiffs' copyright. And Defendants know that these activities infringe. When Plaintiffs' counsel attempted to replicate regurgitations cited in the *Daily News* complaint, using precisely the same methods, ChatGPT refused on copyright grounds. *Id.* ¶ 67.

Again, since the DMCA violation is complete upon removal, it is not necessary that ChatGPT actually did regurgitate or abridge Plaintiffs' articles. But if that were required, it is at least plausible enough that such regurgitation and abridgement occurred to warrant continuing the discovery Plaintiffs have already been conducting into this issue. *See* Section III.B, *supra*.

Inducing, enabling, or facilitating users to infringe. Plaintiffs will also allege that Defendants knew, or had reason know, that removing CMI would induce, enable, or facilitate users to infringe Plaintiffs' copyright by distributing infringing responses to a future audience. Defendants know users distribute ChatGPT responses this way—they even advertise ChatGPT for that purpose. PFAC ¶ 91. Defendants know their users are capable of infringing, including when the users deploy their products as advertised: they have committed to indemnifying their commercial clients in just these circumstances. *Id.* ¶ 94. They also at least have reason to know that removing CMI from training data—causing it not to be included in outputs—would induce

users to infringe. That is because some people respect copyright or fear liability. *Id.* ¶ 92. For example, a user, who possesses a copy, but not the copyright notice, will not know that the work is copyrighted, and will be more likely to make unlawful uses of it than a user who knows that Plaintiffs own copyright to the article and could sue them for distributing it further.

Facilitating Defendants’ training-based infringement. Plaintiffs will also allege that by removing CMI, Defendants facilitate their own training-based infringement. Through training, Defendants aim for ChatGPT to mimic the expression contained in the works on which it was trained. Except when referring to copyrighted material, ordinary English writers do not include CMI—especially copyright notices. But because of the prevalence of CMI in the content on which ChatGPT was trained, including CMI in the training data would have caused it to falsely learn otherwise. Removing CMI at the outset, then, prevented ChatGPT from learning false things about how people use English and facilitated Defendants’ training of ChatGPT. PFAC ¶ 95.

Facilitating Defendants’ copying. Last, Plaintiffs will allege that removing CMI facilitated Defendants’ copying of Plaintiffs’ works. CMI-less works take up less data, so copying them requires fewer computational and storage resources. PFAC ¶ 96. This in turn makes it easier to copy Plaintiffs’ works.

Thus, Plaintiffs alleged scienter in multiple respects. So the proposed amendments will not be futile as to Defendants’ Rule 12(b)(6) motion.

IV. CONCLUSION

The Court should grant Plaintiffs leave to amend their Complaint. In the alternative, it should permit them to continue taking discovery of Defendants' dissemination of their works.

RESPECTFULLY SUBMITTED,

/s/ Stephen Stich Match

Jon Loevy (*pro hac vice*)
Michael Kanovitz (*pro hac vice*)
Lauren Carbajal (*pro hac vice*)
Stephen Stich Match (No. 5567854)
Matthew Topic (*pro hac vice*)
Thomas Kayes (*pro hac vice*)
Steven Art (*pro hac vice*)
Kyle Wallenberg (*pro hac vice*)

LOEVY & LOEVY
311 North Aberdeen, 3rd Floor
Chicago, IL 60607
312-243-5900
jon@loevy.com
mike@loevy.com
carbajal@loevy.com
match@loevy.com
matt@loevy.com
kayes@loevy.com
steve@loevy.com
wallenberg@loevy.com

November 21, 2024