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 OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC, Defendants.	No. 1:24-cv-01514-CM

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

TABLE OF CONTENTS

<u>TABLE OF AUTHORITIES</u>

Page

I.   INTRODUCTION

Lacking any strong argument against amendment, Defendants oppose only by conflating Plaintiffs' current standing theory with one the Court rejected, ignoring inconvenient Second Circuit precedent under *TransUnion*, and neglecting key allegations in the Proposed First Amended Complaint on scienter. They also ignore that, when faced with a largely identical amended complaint in *Intercept*, Judge Rakoff found standing under the DMCA and held that the plaintiff stated a CMI removal claim against OpenAI. Indeed, the only material differences between the amended complaints are that Plaintiffs here allege facts supporting an ***additional*** standing theory—unjust enrichment—and two further grounds for scienter. The Court should reach the same conclusion and allow the proposed amendments.

II.   ARGUMENT

A.   The proposed amendments establish Plaintiffs' standing.

1.   Plaintiffs' injuries are closely analogous to copyright infringement on the specific facts of this case.

Plaintiffs argued that they have alleged standing analogous to copyright infringement because, on the specific facts of this case, Defendants removed their CMI through a technical process that constitutes actual *prima facie* infringement. P. Br. at 3-6, ECF No. 119. In resisting that analogy, Defendants misconstrue this Court's prior opinion, Plaintiffs' standing theory, and the Supreme Court's holding in *TransUnion*.

Defendants cast this Court as holding that dissemination is an absolute prerequisite for a concrete injury. *See* D. Opp. at 7, ECF No. 122 ("[T]he Court has already decided as a matter of law that standing requires dissemination of Plaintiffs' works without CMI."). But the Court held no such thing. It required dissemination only after rejecting the sole alternative proposal on offer:

that violations of the DMCA are *per se* analogous to "injury for interference with property." Order at 6, ECF No. 117 ("Order"). It did not preclude standing theories the parties did not raise.

The Court, moreover, neither considered nor rejected the standing theory Plaintiffs now advance. In opposing Defendants' motion to dismiss, Plaintiffs argued that having one's CMI removed is itself an Article III injury because removal *per se* interferes with property. *See* ECF No. 70 at 6-7. That argument entails standing for **all** plaintiffs who allege Section 1202(b)(1) violations, since CMI removal—hence interference with property—is an element of the claim. *See* 17 U.S.C. § 1202(b)(1). The Court rejected that argument on its terms, holding that "the mere removal of identifying information from a copyrighted work" does not confer standing. Order at 7. But the current theory entails standing for only **some** Section 1202(b)(1) plaintiffs, namely those whose CMI was removed through an act that constitutes *prima facie* copyright infringement. Those plaintiffs have suffered an injury analogous to the historical injury of copyright infringement: their copyright has literally been infringed. The Court has held nothing contrary.[1]

Defendants mount no real resistance to the adequacy of Plaintiffs' allegations that they removed Plaintiffs' CMI through *prima facie* infringement.[2] They argue instead that the allegations are irrelevant because they go to Defendants' conduct rather than Plaintiffs' harm. *See*

---

[1] For this reason, Defendants are wrong to accuse Plaintiffs of seeking "reconsideration in disguise," D. Opp. at 8, as that accusation wrongly conflates Plaintiffs' standing theories. For avoidance of doubt, Plaintiffs are not now asking the Court to reconsider its rejection of their initial standing theory, though they reserve all rights as to that theory in any eventual appeal.

[2] The closest they come is to point out that Plaintiffs have not brought a copyright infringement claim because they have not alleged that they registered their copyrights with the Copyright Office. *See* D. Opp. at 9. But while registering a copyright might be a prerequisite to an infringement suit, it has no effect on the relevant question here: whether OpenAI has *prima facie* infringed Plaintiffs' copyright. *See Palmer/kane LLC v. Gareth Stevens Publ'g*, No. 15-cv-7404, 2016 WL 6238612, at *1 (S.D.N.Y. Oct. 24, 2016) (holding that "registration of a copyright claim is not a condition of copyright protection" but that it is "a prerequisite to bringing a civil copyright infringement action).

D. Opp. at 9. But the allegations go to both, and establish an Article III harm because infringement

has always been a harm in itself. *See* P. Br. at 5 (citing authorities that infringement alone is a

harm). So because infringement itself is an Article III injury, and because Plaintiffs' injuries are

closely analogous to infringement, Plaintiffs have identified the required "close relationship to a

harm traditionally recognized as providing a basis for a lawsuit in American courts." *TransUnion*

LLC v. Ramirez, 594 U.S. 413, 424 (2021) (citation omitted). Just as the statutory violations in

TransUnion sufficed for standing when they also ***resembled*** defamation, so too are the statutory

violations here sufficient when the not only resemble, but ***constitute***, infringement of copyright, a

historically recognized harm. Article III requires no more.

2.   Plaintiffs' injuries are also closely analogous to unjust enrichment.

Plaintiffs argued that Defendants injured them by unlawfully profiting from removing

Plaintiffs' CMI—thus, by realizing profits that rightly belong to Plaintiffs. Further, this injury is

analogous to unjust enrichment, which has a strong common-law pedigree under *TransUnion*, and

which has historically led to disgorgement of profits made in violation of a plaintiff's legally

protected rights—here, rights conferred by the DMCA—whether or not they led to an additional

economic loss. These premises entail standing under *TransUnion*. *See* P. Br. at 6-9.

Defendants do not dispute any of this directly. Instead, they argue that federal courts lack

jurisdiction over claims alleging unlawful profits rightly belonging to the plaintiff unless those

profits come with an additional economic loss. D. Opp. at 12. But they fail to even acknowledge

the post-*TransUnion* Second Circuit case cited in Plaintiffs' opening brief that holds the opposite.

See P. Br. at 6 (citing *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)).

Beyond that, none of their three sub-arguments avails.

First, Defendants argue that Plaintiffs conflate the elements of unjust enrichment—defined by state law—with federal standing law under Article III. *See* D. Opp. at 12. But there is nothing to conflate. After all, historical state-law claims ***determine*** federal standing law. Article III is satisfied when a plaintiff asserts an injury with a "'close relationship' to a harm 'traditionally' recognized as providing a basis for a lawsuit in American courts," and specifically when the plaintiff's injury has a "close historical or common-law analogue." *TransUnion*, 594 U.S. at 424 (quoting *Spokeo, Inc. v. Robins*, 578 U.S. 330, 341 (2016)). And these common-law analogues will typically be state-law claims. *TransUnion* itself is a prime example. Citing the Restatement of Torts, the Court held that the plaintiffs had standing if and only if they satisfied the "publication" element of common-law defamation—a state-law claim. *See id*. at 434. So this argument fails.

Second, Defendants try to liken Plaintiffs' arguments to those made by the dissents in *TransUnion* and *Thole v. U. S. Bank N.A*, 590 U.S. 538 (2020). *See* D. Opp. at 10-12. But neither majority opinion considered the relationship between Article III and unjust enrichment, and Defendants supply no citation to the contrary. *TransUnion* was about an analogy to defamation, and Justice Thomas's dissent mentions unjust enrichment only in passing. *See TransUnion*, 594 U.S. at 459 (Thomas, J., dissenting). In *Thole*, the dissent raised unjust enrichment to argue that the plaintiffs had standing as beneficiaries of a defined benefit plan under ERISA akin to the beneficiary of a private trust. *See Thole*, 590 U.S. at 558-59 (Sotomayor, J., dissenting). But as the dissent observed, the Court rejected that argument because beneficiaries of a defined benefit plan "are not similarly situated to the beneficiaries of a private trust," *id*. at 560 (Sotomayor, J. dissenting) (quoting *Thole*, 590 U.S. at 542), not because Article III bars some actions analogous to unjust enrichment. *Thole* thus has no bearing on this case.

Third, Defendants cite out-of-jurisdiction cases they portray to hold that standing for unjust

enrichment or analogous claims requires an additional economic loss. But at least one of the cases,

Baehr v. Creig Northrop Team, P.C., 953 F.3d 244 (4th Cir. 2020), if anything, holds the opposite.

There the Fourth Circuit rejected an analogy between unjust enrichment and statutory damages for

violations of the Real Estate Settlement Procedures Act ("RESPA"). *See id*. at 257-58. Unlike

unjust enrichment, RESPA's statutory damages do not disgorge the defendant's unlawful gain, but

are instead pegged to the plaintiff's loss. *See* 12 U.S.C. § 2607(d)(2) (providing treble damages).

In light of that difference, the court held, RESPA does not protect against injuries analogous to

unjust enrichment—namely, a defendant's unlawful gain. *See Baehr*, 953 F.3d at 257-58. But it

took pains to note that the defendant's same actions "might give rise to liability in a lawsuit brought

under the unjust enrichment cause of action" rather than RESPA, confirming that Article III

standing does not require an additional economic loss. *Id*. at 257. And the DMCA's statutory

damages provision is analogous to unjust enrichment, not RESPA, since the DMCA's provision is

one of disgorgement. *See* P. Br. at 7-8 (making this argument). So if anything, *Baehr* confirms

Plaintiffs' standing. More, one of Defendants' other cases, *Del Vecchio v. Amazon.com, Inc*., No.

11-cv-366, 2012 WL 1997697 (W.D. Wash. June 1, 2012), does not discuss standing at all.[3]

Defendants' other out-of-jurisdiction cases all predate *TransUnion* and *Spokeo* and thus

did not apply contemporary principles. *See* D. Opp. at 12-13 & n.2 (citing cases from 2015 and

earlier).[4] In particular, they did not consider whether the alleged injuries existed historically or at

---

[3] Defendants cite *Del Vecchio* for the proposition that unjust enrichment (under Washington law)
supposedly requires an additional monetary loss. D. Br. at 13 n.2. But *Del Vecchio* cites no
authority holding that. The Washington case it cites, *Dragt v. Dragt/DeTray, LLC*, 161 P.3d 473,
576 (2007), adopts Section 1 of the Third Restatement, which states that the plaintiff need suffer
no "observable loss." Restatement (Third) of Restitution and Unjust Enrichment, § 1 cmt. a.
[4] One of Defendants' cases, *McNamara v. City of Chicago*, 138 F.3d 1219 (7th Cir. 1998), also
says nothing remotely related to unjust enrichment. The plaintiffs challenged the

common law, or were analogous to such injuries. Plus, they cannot overcome the Second Circuit's application of contemporary principles in *Packer* to find standing based solely on a defendant's unlawful gain without an additional economic loss.

B.   In the alternative, jurisdictional discovery is warranted.

If the Court holds that standing requires dissemination, it should allow discovery into whether OpenAI has disseminated Plaintiffs' articles. As an initial matter, Defendants have no response to the fact that such discovery would have been complete by now had they timely responded to Plaintiffs' RFPs, and for that reason is plainly warranted here. *See* P. Br. at 11-12.

Defendants' responses also do not persuade. They argue that Plaintiffs only "speculat[e]" that ChatGPT might have disseminated their articles, citing the Court's holding that it "seems remote" that ChatGPT plagiarized Plaintiffs' content. D. Opp. at 20 (citing Order at 9). But the Court was applying the standard for jurisdiction over a claim for injunctive relief, which requires a "substantial" risk of imminent harm. Order at 9. That standard is far more stringent than the low bar for jurisdictional discovery. *See, e.g.*, *Ayyash v. Bank Al-Madina*, No. 04-cv-9201, 2006 WL 587342, at *5 (S.D.N.Y. Mar. 9, 2006) (holding jurisdictional discovery warranted where the plaintiff has made a "threshold showing that there is some basis for the assertion of jurisdiction"). Indeed, the discovery standard must be lower than that for alleging jurisdiction, for otherwise every plaintiff entitled to discovery on jurisdiction will have established jurisdiction, and jurisdictional discovery would be an empty vessel. Here, Plaintiffs cannot allege with certainty that ChatGPT disseminated its articles. But the presence of tens of thousands of its articles in the training sets, plus ChatGPT's propensity to plagiarize, at least makes the requisite start. *See* P. Br. at 11.

---

constitutionality of an affirmative action plan used to promote firefighters. The court held that plaintiffs would typically be required to allege that they would have been promoted absent the allegedly unlawful plan. *See id*. at 1221.

Defendants next argue that Plaintiffs do not need discovery because other parties were able to identify disseminations without it. D. Opp. at 20. That argument is disingenuous. Those parties **created** the disseminations but did not identify disseminations produced by ChatGPT in response to prompts by ***others***, which is what Plaintiffs' discovery seeks. PFAC ¶ 77, ECF No. 119-1. While Plaintiffs did not create disseminations, they plausibly alleged that that was due to recent changes OpenAI made to ChatGPT and does not reflect the state of the product years ago, when OpenAI was more cavalier about copyright issues. PFAC ¶ 67.

Defendants last argue that "Plaintiffs' inability to recognize their own injury without discovery only confirms that their purported harm is speculative." D. Opp. at 21. If accepted, that argument would prove that plaintiffs can never get jurisdictional discovery to establish injury in fact. But that is not the law. *See, e.g.*, *Dunne v. Ricolcol*, No. 21-56254, 2024 WL 5088112, at *1 (9th Cir. Dec. 12, 2024) (permitting court on remand to order jurisdictional discovery into injury in fact); *Marsh & McLennan Agency LLC v. Williams*, No. 22-cv-8920, 2023 WL 4534984, at *1 (S.D.N.Y. July 13, 2023) (ordering jurisdictional discovery into injury in fact).

C.   The proposed amendments satisfy Rule 12(b)(6).

Defendants do not dispute that the proposed amendments resolve two of their initial grounds for dismissal: identification of the works at issue, and allegations supporting Defendants' intent to remove CMI. And in *Intercept*, Judge Rakoff resolved the other two issues in the plaintiff's favor based on a materially identical amended complaint. *Compare* Match Decl. Ex. 1 *with* ECF No. 119-1; *see* Match Decl. Ex. 2.[5] This Court should do the same.

---

[5] Because Judge Rakoff has thus far only issued a bottom-line order, it is unclear which scienter theories he accepted. But since he allowed the case to go forward, he accepted at least one. That was necessarily one of the first three theories presented in this brief and Plaintiffs' opening brief, since the *Intercept* plaintiff did not raise the other two.

As to the third issue, statutory standing, OpenAI adds little of substance to its prior briefing. It cites *CoxCom, Inc. v. Chaffee*, 536 F.3d 101 (1st Cir. 2008), which it mistakenly casts as a Second Circuit case, for the proposition that the "question of constitutional standing is not the same as establishing [] standing to sue … as a 'person injured' under the DMCA." D. Opp. at 14 (quoting *CoxCom*, 536 F.3d at 107 n.7). But the First Circuit was simply noting that Congress cannot create standing when Article III forbids it. *See CoxCom*, 536 F.3d at 107 n.7 ("Congress cannot erase Article III's standing requirements by statutorily granting the right to sue to a plaintiff who would not otherwise have standing."). It did not endorse Defendants' proposition—that the bar for injury is higher under the DMCA than the Constitution—since the defendants made no arguments under the DMCA's injury provision at all. *See id*. *CoxCom* also involved claims under a different section of the DMCA—Section 1201—and so has no bearing here. *See id*. at 104.

On the last issue, the second scienter element, Defendants' responses all fail.

Concealing Defendants' training-based infringement from their users. Defendants' only response is to misconstrue the theory. Defendants make the banal observation that removing CMI cannot conceal "a non-public dataset." D. Opp. at 15. But Plaintiffs have not alleged otherwise. The point, rather, is that Defendants concealed their own infringement via the process of reproducing Plaintiffs' articles to ChatGPT users. A reproduction that incorporates Plaintiffs' works **and** Plaintiffs' CMI would communicate to users that the reproduction resulted from an infringing copy in the training set, and thus that OpenAI committed training-based infringement. Conversely, one that incorporates Plaintiffs' works **without** Plaintiffs' CMI conveys no such thing. So omitting CMI from the reproduction works a concealment. And critically, Plaintiffs alleged,

removing CMI from the training set causes it to be absent from the reproduction. *See* PFAC ¶¶ 81, 89. Defendants have no direct response to this theory properly construed.[6]

 Concealing Defendants' output-based infringement from their users. Defendants at least describe this theory mostly correctly,[7] yet their rebuttals are wrong. They argue that their knowledge of the "general phenomenon of regurgitation" does not equate to knowledge that removing CMI will conceal an infringement. D. Opp. at 16-17. But Defendants' reason to know this derives not only from their expertise in regurgitation, but the fact that OpenAI designed the product at issue, and that removing CMI in training causes CMI to be absent from ChatGPT outputs—clearly a plausible inference at the pleading stage given the low bar for scienter. *See* P. Br. at 14-15 (collecting cases on scienter pleading standards). Defendants also say that Plaintiffs fail to allege this causal connection. D. Opp. at 17. But that is just false. *See* PFAC ¶ 81 ("If ChatGPT was trained on works of journalism that included the original author, title, and copyright information, ChatGPT would have learned to communicate that information."). And while Defendants point out that Plaintiffs have not alleged dissemination of their articles, D. Opp. at 17, they fail to contend with Plaintiffs' argument that this is unnecessary under the DMCA, which does not require an infringement to have occurred. *See* P. Br. at 15-16 (citing *Murphy v. Millennium Radio Grp. LLC*, No. 08-cv-1743, 2015 WL 419884, at *4-5 (D.N.J. Jan. 30, 2015)).

 Inducing, enabling, or facilitating users to infringe. Apart from objections already addressed above, Defendants dispute this theory by assailing as "wholly conclusory" the

---

[6] Defendants' citation to *Tremblay v. OpenAI, Inc.*, 716 F. Supp. 3d 772 (N.D. Cal. 2024) confounds. The plaintiffs there argued that removing CMI would enable users to infringe, which is completely different from the issue Defendants cite it for. *See id*. at 778.

[7] Defendants suggest Plaintiffs to be alleging that Defendants removed the CMI with a purpose to conceal infringement. *See* D. Opp. at 16. But the second scienter element requires only knowledge, not purpose. *See* 17 U.S.C. § 1202(b).

proposition that users would refrain from distributing works they know to be copyrighted. D. Opp. at 17-18. But it is common sense—and certainly plausible at the pleading stage—that people are less likely to perform conduct they know is illegal.

Facilitating Defendants' training-based infringement. Defendants do not dispute the adequacy of Plaintiffs' allegation that removing CMI facilitates LLM training. They instead argue that Plaintiffs have not shown how LLM training constitutes infringement. D. Opp. at 18-19. They suggest that the infringement must be one of several discrete parts of the training process, such as the act of downloading Plaintiffs' works. But that defines the infringement too narrowly. Training itself is the infringement, as Defendants have recognized in other cases. *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 (defining the "genuinely important issue at the heart of this lawsuit" as "whether it is fair use under copyright law to use publicly accessible content to train generative models").

Facilitating Defendants' copying. Defendants complain that Plaintiffs supposedly have not explained why removing CMI facilitates copying, D. Opp. at 19, but that is false. Articles without CMI take up fewer computational and storage resources, which frees up those resources for further copying of other articles. PFAC ¶ 96. And *contra* Defendants, this theory does not contradict Plaintiffs' allegation that Defendants first download an article and then extract CMI. D. Opp. at 19 (citing PFAC ¶ 46). Removing CMI from already-downloaded articles saves resources compared to leaving the CMI on them.

III.   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.

- 11 -

RESPECTFULLY SUBMITTED,

/s/ Stephen Stich Match

Jon Loevy (*pro hac vice*)
Michael Kanovitz (*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
match@loevy.com
matt@loevy.com
kayes@loevy.com
steve@loevy.com
wallenberg@loevy.com

January 21, 2025