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-SHS

NOTICE OF MOTION

PLEASE TAKE NOTICE THAT upon the annexed Memorandum of Law, the Declaration of Stephen Stich Match, and the exhibits thereto, the undersigned hereby moves this Court on behalf of Plaintiffs Raw Story Media, Inc., and AlterNet Media, Inc. for an order granting reconsideration of the orders granting Defendants' motion to dismiss and denying Plaintiffs leave to amend their complaint, ECF Nos. 117, 137, and permitting Plaintiffs to file an amended complaint within seven days of the Court's order granting this motion.

RESPECTFULLY SUBMITTED,

/s/ Stephen Stich Match

Jon Loevy (*pro hac vice*)
Michael Kanovitz (*pro hac vice*)
Matthew Topic (*pro hac vice*)
Stephen Stich Match (No. 5567854)
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
matt@loevy.com
match@loevy.com
kayes@loevy.com
steve@loevy.com
wallenberg@loevy.com

April 18, 2025

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-SHS

MEMORANDUM OF LAW IN SUPPORT OF *RAW STORY* PLAINTIFFS' MOTION FOR RECONSIDERATION OF DISMISSAL AND DENIAL OF LEAVE TO AMEND COMPLAINT

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page

Cases

Other Authorities

I. INTRODUCTION

In this multidistrict litigation, all but one set of plaintiffs have survived challenges to their standing to bring copyright management removal claims against OpenAI under section 1202(b)(1) of the Digital Millennium Copyright Act. The outliers, Raw Story Media, Inc. and AlterNet, Media, Inc. (the "*Raw Story* Plaintiffs") respectfully seek reconsideration of the dismissal of their complaint, and denial of leave to amend it to bring it in line with claims this Court has already upheld against dismissal. Doing so will achieve the consistency in pretrial rulings and discovery efficiencies that motivated OpenAI to seek an MDL in the first place, and ensure that the *Raw Story* plaintiffs, like the others, have their day in court to challenge OpenAI's violations of the DMCA in creating their massively lucrative AI chatbot.

II. BACKGROUND

A. The *Raw Story* litigation.

The *Raw Story* Plaintiffs brought a single-count complaint against OpenAI for removing their copyright management information in violation of 17 U.S.C. § 1202(b)(1).[1] Compl. *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Feb. 28, 2024), ECF No. 1. OpenAI moved to dismiss for lack of Article III standing and failure to state a claim. *See* Memorandum of Law in Support of Defendants' Motion to Dismiss at 4-18, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Apr. 29, 2024), ECF No. 68.[2] The *Raw Story* Plaintiffs opposed, arguing that removal of copyright management information presents a sufficiently close analogy to copyright infringement to ground standing under *TransUnion LLC v. Ramirez*, 594 U.S. 413 (2021). *See* Plaintiffs' Memorandum of Law in Opposition to Defendants' Motion to Dismiss at

[1] Unlike the other parties in this MDL that brought DMCA claims, the *Raw Story* Plaintiffs did not sue Microsoft and did not bring any claims under section 1202(b)(3).
[2] OpenAI raised what it has sometimes called "statutory standing" under the failure-to-state-a-claim rubric. *See id*. at 11-12.

5-11, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. May 13, 2024), ECF No. 71. They also argued that the complaint stated a claim. *See id*. at 10-20. They further sought leave to replead should the Court dismiss the complaint. *See id*. at 21.

In a November 7, 2024 order, the court granted OpenAI's motion on standing grounds and did not reach the adequacy of the pleadings. *See* Match Decl. Ex. 1 ("MTD Order"). It was "not convinced that injury for interference with property provides the necessary 'close historical or common-law analogue' to Plaintiffs' alleged injury," and thus rejected the argument that the injury "bears a 'close relationship' to the tort of copyright infringement." *Id*. at 6. The court also held that removal of CMI does not constitute a concrete injury "absent dissemination." *Id*. at 7. And it denied leave to amend, but without prejudice to the filing of a motion that attached the proposed pleading "together with an explanation of why the proposed amendment would not be futile." *Id*. at 10.

The *Raw Story* Plaintiffs filed that motion two weeks later. The Proposed First Amended Complaint ("PFAC") that accompanied it was substantively identical to the section 1202(b)(1)-related allegations in The Center for Investigative Reporting's First Amended Complaint, which was drafted by the same counsel. *Compare* Match Decl. Ex. 2 (the PFAC) *with* First Amended Complaint, *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-cv-04872 (Sept. 24, 2024), ECF No. 88 (the "CIR FAC"). For that reason, the briefing on the futility of the amendments closely tracked the briefing in *CIR* pertaining to OpenAI's motion to dismiss the section 1202(b)(1) claim. The *Raw Story* Plaintiffs continued to argue that CMI removal presented a sufficiently close analogy to copyright infringement under *TransUnion*, particularly because OpenAI allegedly removed the CMI through actual *prima facie* copyright infringement. *See* Plaintiffs' Memorandum of Law in Support of Their Motion for Leave to Amend Complaint or, in

the Alternative, to Continue Taking Jurisdictional Discovery at 3-6, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Nov. 21, 2024), ECF No. 119 ("MTA"). The *Raw Story* Plaintiffs also expanded on an argument made in *CIR* that removing CMI is analogous to the common-law action for unjust enrichment, citing Second Circuit precedent that disgorgement of ill-gotten profits can ground standing after *TransUnion* even without a separate economic loss, and arguing that OpenAI unjustly enriched itself by removing their CMI. *See id*. at 6-9; *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, 145 S. Ct. 550 (2024); *see also* Plaintiff's Combined Response to Defendants' Motions to Dismiss the First Amended Complaint at 20-21, *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-cv-04872 (S.D.N.Y. Nov. 5, 2024), ECF No. 133.

On the merits, the dispute centered on whether the *Raw Story* had statutory standing as a "person injured" under 17 U.S.C. § 1203(a) and whether the proposed new pleading plausibly alleged scienter. *See* MTA at 12-17; OpenAI's Memorandum of Law in Opposition to Plaintiffs' Motion for Leave to Amend Complaint or, in the Alternative, to Continue Taking Jurisdictional Discovery at 14-19, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (Dec. 20, 2024), ECF No. 122 ("MTA Opp."); Plaintiffs' Reply in Support of Their Motion for Leave to Amend Complaint or, in the Alternative, to Continue Taking Jurisdictionatl Discovery at 7-10, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Jan. 21, 2025), ECF No. 127.

The court denied leave to amend on April 3, 2025. It held again that the "unauthorized removal of CMI from … copyrighted-protected work" is not a concrete injury absent dissemination, and that copyright infringement is not a close historical analogue to CMI removal. Match Decl. Ex. 3 at 2-3 ("MTA Order"). It also rejected the analogy to unjust enrichment because the complaint did not explain how OpenAI's profits "would have found its way to or from

Plaintiffs' pockets," without discussing the Second Circuit's holding in *Packer*. *Id*. at 3. Because it ruled on Article III standing, the court did not address 17 U.S.C. § 1203(a) or scienter.

The court had earlier allowed discovery to proceed pending OpenAI's motion to dismiss. Match Decl. ¶ 8. Thus, the parties pursued written discovery until the court dismissed the initial complaint in November. *Id*. ¶ 9. So the case is in roughly the same discovery posture as several of the other cases that have been centralized, and is even ahead in some respects. For example, OpenAI has provided the *Raw Story* Plaintiffs with copies of their works as they are contained in OpenAI's training sets, which it has not done for CIR even though the latter requested similar documents in October. *Id*. ¶¶ 10-11.

B. The MDL proceeding.

After the court dismissed the *Raw Story* complaint, OpenAI asked the Judicial Panel on Multidistrict Litigation to transfer all the copyright-adjacent cases against it, including *Raw Story*, into a multidistrict litigation ("MDL"). *See* Brief in Support of OpenAI's Motion for Transfer of Actions Pursuant to 28 U.S.C. § 1407 for Coordinated or Consolidated Pretrial Proceedings, *In re OpenAI, Inc. Copyright Infringement Litigation*, MDL No. 3143 (J.P.M.L. Dec. 6, 2024), ECF No. 1-1 ("MDL Motion"). OpenAI argued at length that an MDL was necessary to, "avoid inconsistent rulings on important and emerging questions of law," warning that any such inconsistencies "will not only hamper and prejudice the parties in this litigation, but risk creating confusion and uncertainty in the development of LLMs and artificial intelligence more broadly, stymying growth and innovation in a burgeoning field." *Id*. at 16.

In support, OpenAI highlighted the diverging conclusions reached in *Raw Story* and *Intercept*, which had found standing for CMI removal claims against OpenAI. *See The Intercept Media, Inc. v. OpenAI, Inc*., No. 24-cv-01515, 2025 WL 556019, at *4-6 (S.D.N.Y. Feb. 20, 2025). Specifically, OpenAI argued:

> [T]wo judges have reached different outcomes on the same legal issue presented by two near-identical complaints brought by the same counsel, in the same district. Section 1407 centralization is intended to prevent precisely such inconsistencies. These fractures will only grow as the litigation progresses. This Panel can ensure that the courts issue consistent guidance in this vanguard litigation by centralizing the litigation in a single district, before a single judge.

Id. at 17.[3] It doubled down on reply, stressing that the risk of inconsistent rulings is "among the most compelling bases for centralization." Reply Brief in Support of OpenAI's Motion for Transfer of Actions Pursuant to 28 U.S.C. § 1407 for Coordinated or Consolidated Pretrial Proceedings at 4, *In re OpenAI, Inc. Copyright Infringement Litigation*, MDL No. 3143 (J.P.M.L. Jan. 10, 2025), ECF No. 62.

OpenAI also argued that centralization "is necessary to limit duplicative discovery." MDL Motion at 14. In particular, OpenAI complained that without centralization, its witnesses "are virtually guaranteed to sit for multiple depositions," *id*. at 15, and criticized the plaintiff groups for purportedly not agreeing "to single depositions of witnesses," *id*.

The Panel granted OpenAI its wish for a coordinated and centralized proceeding. It reasoned that "[c]entralization will allow a single judge to … eliminate inconsistent rulings." Transfer Order at 3, *In re OpenAI, Inc. Copyright Infringement Litigation*, MDL No. 3143 (J.P.M.L. Apr. 3, 2025), ECF No. 85. It also held that centralization will "eliminate duplicative

[3] In opposing centralization, the *Raw Story* Plaintiffs, together with The Intercept, argued that OpenAI had not shown a direct conflict between the *Raw Story* and *Intercept* decisions. Though *Raw Story* and *Intercept* started from similar complaints, the *Intercept* court had allowed The Intercept to file an amended complaint, which added allegations The Intercept argued bolstered its argument for standing. When these parties responded to OpenAI's motion, the *Intercept* court had issued a bottom-line order but not a full opinion, and it was therefore possible that those allegations had affected the outcome. *See* DMCA Plaintiffs' Response to OpenAI's Motion for Transfer of Actions Pursuant to 28 U.S.C. § 1407 for Coordinated or Consolidated Pretrial Proceedings at 4, 12, *In re Open, Inc. Copyright Infringement Litigation*, MDL No. 3143 (J.P.M.L. Jan. 3, 2025), ECF No. 51. Now that the court has issued a full opinion, it does appear that *Intercept* and *Raw Story* cannot be reconciled. So too for *Raw Story* and *New York Times*.

depositions of defense witnesses." *Id*. The Panel issued its order on the same day as the *Raw Story* order denying leave to amend, but after that order had issued.

III. ARGUMENT

A. The MDL order constitutes a changed circumstance warranting reconsideration.

Under Fed. R. Civ. P. 54(b), a court may "reconsider a decision or order at any time before the entry of final judgment." *Commerzbank AG v. U.S. Bank, N.A.*, 100 F.4th 362 (2d Cir.), cert. denied, 145 S. Ct. 279 (2024). Though the court disposed of all open substantive motions prior to the MDL order, final judgment has not yet entered. *See* Fed. R. Civ. P. 58(a) ("Every judgment and amended judgment must be set out in a separate document."); Docket, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y.) (lacking reference to a separate document entering final judgment). Thus, Rule 54(b) governs this motion. In the alternative, the *Raw Story* Plaintiffs seek reconsideration under Fed. R. Civ. P. 60(b)(6), which allows for relief from a final judgment or order for "any other reason that justifies relief." Under both rules, courts must balance interests in finality and judicial resources with other competing considerations. *See Zdanok v. Glidden Co., Durkee Famous Foods Div.*, 327 F.2d 944, 953 (2d Cir. 1964) ("[W]here litigants have once battled for the court's decision, they should neither be required, nor without good reason permitted, to battle for it again."); *In re Terrorist Attacks on Sept. 11, 2001*, 741 F.3d 353, 357 (2d Cir. 2013) ("Properly applied Rule 60(b) strikes a balance between serving the ends of justice and preserving the finality of judgments.").

Transferee courts have the power to override transferor courts' prior rulings. *See In re Grand Jury Proc. (Kluger)*, 827 F.2d 868, 871 n.3 (2d Cir. 1987) ("A transfer under 28 U.S.C. § 1407 transfers the action lock, stock, and barrel. The transferee district court has the power and the obligation to modify or rescind any orders in effect in the transferred case which it concludes

are incorrect.") (cleaned up). In the MDL context, finality often gives way to the interests that motivated the transfer in the first place: consistency. Thus, in upholding a district court's reconsideration of a dismissal order, the Fourth Circuit observed that the "constraint" ordinarily warranted when a district court is asked to reconsider a prior order "is not always justified in the multidistrict litigation context, where there is a need for consistent treatment of consolidated cases." *Pinney v. Nokia, Inc*., 402 F.3d 430, 453 (4th Cir. 2005). For the same reasons, other MDL courts adjudicating motions to reconsider dismissal orders have found that the need to ensure consistent pretrial rulings "can constitute, as here, the type of 'significant' and 'fundamental' change warranting reconsideration of the order by the transferee court." *Degulis v. LXR Biotechnology, Inc*., 928 F. Supp. 1301, 1309 (S.D.N.Y. 1996) (quoting *In re Long Distance Telecomm. Litig*., 612 F. Supp. 892, 903 (E.D. Mich. 1985), aff'd in part, rev'd in part on other grounds, 831 F.2d 627 (6th Cir. 1987)).

Reconsideration is appropriate here for two reasons. First, as in the cases cited above, the transferor courts have decided motions presenting identical issues. This Court, and the court in *Intercept*, held that CMI removal constitutes a concrete injury to property rights analogous to copyright infringement, and rejected OpenAI's argument that standing also requires dissemination. *See New York Times Co. v. Microsoft Corp*., No. 23-CV-11195, 2025 WL 1009179, at *12-15 & n.6 (S.D.N.Y. Apr. 4, 2025); *Intercept*, 2025 WL 556019, at *3-6. The *Raw Story* court reached the opposite conclusion. *See* MTD Order at 6-7. Thus, reconsideration is necessary to correct an inconsistency among pretrial rulings in cases that OpenAI chose to insert into its MDL proposal.

Second, reconsideration will also create efficiencies in discovery—OpenAI's other main reason for seeking an MDL. Though the case has been dismissed since November, the parties have already engaged in substantial discovery, and the *Raw Story* Plaintiffs anticipate being able

to participate without altering the schedule, assuming this motion is decided reasonably promptly. *See* Match Decl. ¶¶ 8-10. On the other hand, if they are forced to appeal, and the Second Circuit joins this Court and *Intercept* in finding standing, then the MDL cases might well be over, and OpenAI would have to do the rest of discovery all over again. Notably, the *Raw Story* Plaintiffs would then have the opportunity to depose OpenAI's witnesses—directly contrary to OpenAI's desire to have its witnessed deposed only once.

OpenAI cannot deny this. It moved for, and obtained, an MDL for precisely these reasons. Indeed, OpenAI's success in obtaining an MDL for these purposes judicially estops it from objecting to these benefits of reconsideration. *See New Hampshire v. Maine*, 532 U.S. 742, 749 (2001) ("Where a party assumes a certain position in a legal proceeding, and succeeds in maintaining that position, he may not thereafter, simply because his interests have changed, assume a contrary position, especially if it be to the prejudice of the party who has acquiesced in the position formerly taken by him.") (cleaned up).

The same result follows under Rule 60(b)(6). For instance, in *In re Terrorist Attacks*, the Second Circuit reversed the district court's denial of a Rule 60(b)(6) motion in an MDL case. The court had upheld the dismissal of claims brought by a first set of plaintiffs, but reached the opposition conclusion in a later opinion involving similarly situated plaintiffs joined later in the MDL process. *See* 741 F.3d at 355-56. The first set of plaintiffs, whose claims had been dismissed, then sought relief under Rule 60(b)(6). The court held that it would be unfair to allow "inconsistent results between two sets of plaintiffs suing for damages based on the same incident," *id*. at 357 and found that result "particularly troubling" because that inconsistency resulted from the timing of the MDL centralization, which had been ordered to "prevent inconsistent pretrial rulings," *id*. at 358.

Similar considerations apply here. The *Raw Story* Plaintiffs' claims arise out of the same actions by OpenAI as those that have been allowed to proceed: OpenAI's removal of CMI from their articles in the same or similar training sets, such as *e.g.*, WebText, WebText2, and sets derived from Common Crawl. *See* PFAC ¶¶ 36-62; *New York Times*, 2025 WL 1009179, at *4, 16-17. So it would be unjust to treat them differently, particularly since it is OpenAI that sought to join them in the MDL. And the inconsistency arose only because OpenAI waited to move for an MDL until after the *Raw Story* and *Intercept* courts had issued conflicting decisions on standing and the briefing in *New York Times*, *Daily News*, and *CIR* was already complete; had it sought the MDL sooner, this Court would have addressed the *Raw Story* Plaintiffs' standing together with the other cases and reached consistent results across the cases.

The Court should therefore reconsider the *Raw Story* court's dismissal of the complaint and denial of the *Raw Story* Plaintiffs' motion for leave to amend their complaint.

B. The *Raw Story* Plaintiffs have standing.

On reconsideration, the Court should hold that the *Raw Story* Plaintiffs have standing.[4] This point need not be belabored, since this Court has already found standing from virtually identical allegations in *New York Times*, *Daily News*, and *CIR*. Like those other plaintiffs, the *Raw Story* Plaintiffs alleged that OpenAI downloaded tens of thousands of their articles and intentionally removed CMI from them before using the CMI-less copies to train its GPT models. *See* PFAC ¶¶ 36-62. Like CIR, the *Raw Story* Plaintiffs also alleged that OpenAI did so through a process that involved creating unauthorized copies and thus constituted actual *prima facie* copyright infringement. *Compare* PFAC ¶¶ 42-52 *with* CIR FAC ¶¶ 59-69. And again like CIR, the *Raw Story* Plaintiffs alleged that OpenAI did so knowing, or having reason to know, that

[4] The *Raw Story* Plaintiffs cite their Proposed First Amended Complaint because, as discussed more fully in Section III.C, *infra*, the Court should allow them to file the amended version.

removing CMI would conceal, induce, enable, and facilitate infringement by both OpenAI and its users. *Compare* PFAC ¶¶ 84-96 *with* CIR FAC ¶¶ 108-119. The Court has already found that these injuries are analogous to copyright infringement because "the harm involves an injury to 'another's property right in his original work of authorship'" and that they therefore ground standing. *New York Times*, 2025 WL 1009179, at 14 (quoting *Intercept*, 2025 WL 556019, at *5). It should reach the same conclusion here.[5]

C. The Court should allow the *Raw Story* Plaintiffs to amend their complaint.

Because the *Raw Story* court found no Article III standing, it had no occasion to consider the other reasons why OpenAI argued that the case should not proceed: statutory standing under 17 U.S.C. § 1203(a) and the supposed failure to plead necessary elements of a section 1202(b)(1) claim.

This Court should consider that question vis-à-vis the *Raw Story* Plaintiffs' Proposed First Amended Complaint, not the initial version. Leave to amend is freely given when justice requires, *see* Fed. R. Civ. P. 15(a)(2), and at no point did either the *Raw Story* court or OpenAI give a reason for denying leave other than the supposed inadequacy of the pleadings. For instance, OpenAI did not argue that allowing an amendment would cause it any prejudice, and none is apparent. *See generally* MTA Opp. Further, the *Raw Story* Plaintiffs timely moved for leave to amend a mere two weeks after the court dismissed the complaint and are filing this motion within two weeks of both the order denying leave to amend and the order granting the MDL. Last, this Court granted OpenAI's motion to dismiss the New York Times' section 1202(b)(1) claim without prejudice,

[5] The Court can also find standing under an analogy to unjust enrichment for the reasons given in the *Raw Story*'s Motion to Amend briefing. The *Raw Story* Plaintiffs add only that the court's reason for rejecting that analogy—that "Plaintiffs do not explain how either the benefit or the expense of those costs would have found its way to or from Plaintiffs' pockets," MTA Order at 3—does not account for the Second Circuit's finding of standing in a case involving profits that would not have made their way into the plaintiffs pockets. *See Packer*, 105 F.4th at 51-56.

which has permitted the Times to file a second amended complaint to bolster its allegations under that statute. *See New York Times*, 2025 WL 1009179, at *2. The *Raw Story* Plaintiffs would be doing the same, but for the first time.

The *Raw Story* Plaintiffs' proposed pleading easily survives the objections the *Raw Story* court did not reach, essentially for reasons this Court has already given. As to statutory standing, the Court assumed without deciding that section 1203(a) requires allegations of injury beyond a mere statutory violation but found that the plaintiffs alleged various downstream injuries, such as the concealment of OpenAI's own copyright infringement and enabling and facilitating the copyright infringement of end users. *New York Times*, 2025 WL 1009179, at *15. The proposed pleading alleges just that. *See* PFAC ¶¶ 84-96.[6]

As to the first scienter element, intentional removal, this Court upheld the sufficiency of CIR's complaint based on its allegations about the functioning of the Dragnet and Newspaper algorithms and OpenAI's knowledge about them. *See New York Times*, 2025 WL 1009179, at *15. The *Raw Story* Plaintiffs' proposed pleading, drafted by the same counsel, contains substantively identical allegations and thus suffices. *Compare* PFAC ¶¶ 42-52, 83, 102; *with* CIR FAC ¶¶ 59-69, 103, 125.

As to the second scienter element, actual or constructive knowledge that removing CMI will induce, enable, facilitate, or conceal an infringement, the proposed pleading again suffices. This Court pointed to allegations in the CIR complaint that OpenAI has publicly acknowledged that it uses copyrighted works to train its models and that its models regurgitate content. *See New*

[6] The *Raw Story* Plaintiffs dispute that section 1203(a) imposes any additional requirement. The legislative history makes clear that this section was included merely to create a cause of action for violations of the statute's substantive provisions. *See* S. Rep. 105-190, at 38 (1998) (noting that this section "sets forth the general proposition that civil remedies are available for violations of sections 1201 and 1202").

York Times, 2025 WL 1009179, at * 17. It also relied on CIR's allegations that OpenAI's indemnification policies, and its recent adjustment to ChatGPT's settings to limit regurgitations, show that it was aware of the prospect of end user infringement. *See id*. Again, the *Raw Story* Plaintiffs' proposed pleading contains all these allegations and thus states a claim. Compare PFAC ¶¶ 67, 94 *with* CIR FAC ¶¶ 84, 118.

IV. CONCLUSION

The Court should reconsider the dismissal of the *Raw Story* Plaintiffs' complaint and allow them to file their Proposed First Amended Complaint.

RESPECTFULLY SUBMITTED,

/s/ Stephen Stich Match

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

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

April 18, 2025

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-SHS

DECLARATION OF STEPHEN STICH MATCH

1. My name is Stephen Stich Match.

2. I am over the age of 18 and competent to provide this declaration.

3. The following statements are made on my personal knowledge, and, if sworn as a witness, I can testify competently thereto.

4. I am an attorney at Loevy + Loevy admitted to the state bars of California, Illinois, Massachusetts, and New York. I represent Raw Story Media, Inc. AlterNet Media, Inc., The Intercept Media, Inc., and The Center for Investigative Reporting, Inc.

5. A true and correct copy of the November 7, 2024 order dismissing the complaint and denying leave to amend without prejudice in *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Nov. 7, 2024), ECF No. 117 is attached as Exhibit 1.

6. A true and correct copy of the Proposed First Amended Complaint attached to the *Raw Story* Plaintiffs' Motion for Leave to Amend or, in the Alternative, to Continue Taking

Jurisdictional Discovery, *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Nov. 21, 2024), ECF No. 119-1 is attached as Exhibit 2. The complaint's exhibits are omitted.

7. A true and correct copy of the April 3, 2025 order denying leave to amend in *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01514 (S.D.N.Y. Apr. 3, 2025), ECF No. 137 is attached as Exhibit 3.

8. The *Raw Story* court did not stay discovery pending OpenAI's motion to dismiss.

9. The parties in *Raw Story* engaged in written discovery until the court dismissed the complaint.

10. In discovery, OpenAI provided the *Raw Story* Plaintiffs with copies of their works as they are contained in OpenAI's training sets.

11. On October 11, 2024, The Center for Investigative Reporting served OpenAI with a request for production for copies of its works as they are contained in OpenAI's training sets. As of this date, OpenAI has not produced those documents.

Pursuant to 28 U.S.C. § 1746, I declare under penalty of perjury that the foregoing is true and correct. Executed on April 18, 2025.

/s/ Stephen Stich Match
Stephen Stich Match

EXHIBIT 1

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

x

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

Plaintiffs,

vs.

24 Civ. 01514

OPENAI, INC., OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, and OPENAI HOLDINGS, LLC,

Defendants.

x

DECISION AND ORDER

McMahon, J.:

Plaintiffs Raw Story Media, Inc. and AlterNet Media, Inc. (collectively "Plaintiffs") brought this action, pursuant to the Digital Millennium Copyright Act (the "DMCA"), 17 U.S.C. § 1201 *et seq.*, against OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI Opco LLC, OpenAI Global LLC, and OpenAI Holdings, LLC (collectively, "Defendants" or "OpenAI"). OpenAI seeks to dismiss Plaintiffs' complaint in its entirely, pursuant to Fed. R. Civ. P. 12(b)(1) and 12(b)(6). Dkt. No. 68 ("MTD").

For the reasons below, OpenAI's motion to dismiss is GRANTED. Plaintiffs' motion for leave to replead is DENIED WITHOUT PREJUDICE to renewal on a proper record—which means filing a notice of motion to which a proposed amended pleading is attached, together with an explanation of why the proposed amendment would not be futile.

BACKGROUND

Plaintiffs Raw Story Media, Inc. and AlterNet Media, Inc. are news organizations that have published, collectively, "more than 400,000 breaking news features, investigative news articles and opinion columns" online. Dkt. No. 1, ¶ 14 ("Compl."). Defendants OpenAI are seven interrelated organizations, doing business within the State of New York, responsible for an AI service known as ChatGPT. Compl. ¶¶ 15-22, 24. ChatGPT is an AI-powered large language model ("LLM") that allows paying users to enter text prompts to which ChatGPT will generate responses. According to Plaintiffs, ChatGPT "gives the impression that it is an all-knowing 'intelligent' source of the information being provided." Compl. ¶ 40. However, "ChatGPT does not have any independent knowledge of the information provided in its responses." Compl. ¶ 36. Rather, ChatGPT is trained on large amounts of text, known as "training sets." Compl. ¶ 29. "These training sets range from collections of links posted on the website Reddit to a scrape of most of the internet." Compl. ¶ 29.

Plaintiffs allege that "thousands" of their copyrighted works of journalism were caught in this "scrape," stripped of their author, title, and copyright information, and input into at least three of OpenAI's training sets (WebText, WebText2, and Common Crawl). Compl. ¶¶ 29, 30, 37. Plaintiffs allege that these three training sets were then used to train ChatGPT. Compl. ¶ 29. Since ChatGPT was not provided with the author, title, and copyright information, Plaintiffs allege that ChatGPT would not have learned to communicate that information when fashioning responses to user inquiries that are based on their copyrighted works, Compl. ¶ 38, and that indeed ChatGPT "generally does not provide the author, title, and copyright information applicable to the works on which its responses are based." Compl. ¶ 39. Plaintiffs allege that Defendants' removal of copyright management information ("CMI") from Plaintiffs' works—prior to training ChatGPT—

is a violation of Section 1202(b)(i) of the Digital Millenium Copyright Act (the "DMCA"), for which Plaintiffs are entitled to actual or statutory damages.

Plaintiffs further seek injunctive relief against Defendants. Plaintiffs allege that earlier versions of ChatGPT generated significant amounts of plagiarized content. Compl. ¶ 5. If Plaintiffs' works remain in ChatGPT's repository without any CMI, Plaintiffs allege that there is a substantial likelihood that the current version of ChatGPT will reproduce, verbatim or nearly verbatim, Plaintiffs' copyrighted works without providing the author, title, or copyright information contained in those works. *See* Prayer for Relief.

OpenAI has moved to dismiss the complaint. Dkt. No. 68 ("MTD"). Plaintiff opposes the motion. Dkt. No. 70 ("Opp.").

LEGAL STANDARDS

I. Section 1202(b)(i) of the DMCA

Section 1202(b)(i) of the DMCA provides that:

> No person shall, without the authority of the copyright owner or the law,… intentionally remove or alter any [CMI]…knowing, or with respect to civil remedies under section 1203, having reasonable grounds to know that it will induce enable, facilitate or conceal an infringement of any right under this title.

17 U.S.C. §1202(b). The Second Circuit has found this statute to incorporate a so-called double-scienter requirement: that the defendant both know "that CMI has been removed or altered without authority of the copyright owner or the law" and know, "or hav[e] reasonable grounds to know that such distribution will induce, enable, facilitate, or conceal an infringement." *Mango v. Buzzfeed*, 970 F.3d 167, 171 (2d Cir. 2020) (cleaned up) (internal citations omitted).

II. Article III Standing

"Article III standing requires a concrete injury even in the context of a statutory violation." *Spokeo, Inc. v. Robins,* 578 U.S. 330, 341 (2016). "[T]o establish standing, a plaintiff must show (i) that he suffered an injury in fact that is concrete, particularized, and actual or imminent; (ii) that the injury was likely caused by the defendant, and (iii) that the injury would likely be redressed by judicial relief." *TransUnion LLC v. Ramirez,* 594 U.S. 413, 423 (2021). "The party invoking federal jurisdiction bears the burden of establishing these elements." *Lujan v. Defenders of Wildlife,* 504 U.S. 555, 561 (1992).

"Where, as here, a case is at the pleading stage, the plaintiff must clearly…allege facts demonstrating each element." *Spokeo, Inc. v. Robins,* 578 U.S. 330, 338 (2016) (internal citation and quotations omitted). "[S]tanding is not dispensed in gross; rather, plaintiffs must demonstrate standing…for each form of relief that they seek (for example, injunctive relief and damages)." *TransUnion LLC*, 594 U.S. at 431.

"What makes a harm concrete for the purposes of Article III?" *TransUnion,* 594 U.S. at 424:

> As a general matter, the Court has explained that history and tradition offer a meaningful guide to the types of cases that Article III empowers federal courts to consider. And with respect to the concrete-harm requirement in particular, this Court's opinion in *Spokeo v. Robins* indicated that courts should assess whether the alleged injury to the plaintiff has a close relationship to a harm traditionally recognized as providing a basis for a lawsuit in American courts. That inquiry asks whether plaintiffs have identified a close historical or common-law analogue for their asserted injury. *Spokeo* does not require an exact duplicate in American history and tradition. But *Spokeo* is not an open-ended invitation for federal courts to loosen Article III based on contemporary, evolving beliefs about what kinds of suits should be heard in federal courts.

Id. at 424-25. (internal quotations and citations omitted).

DISCUSSION

Defendants seek to dismiss Plaintiffs' complaint in its entirely. Defendants argue that Plaintiffs lack Article III standing to assert their claims, and that the Court therefore lacks subject-matter jurisdiction over this action pursuant to Fed. R. Civ. P. 12(b)(1). Defendants argue, in the alternative, that Plaintiffs have failed to state a claim upon which relief can be granted, pursuant to Fed. R. Civ. P. 12(b)(6).

"Whether a claimant has standing is 'the threshold question in every federal case, determining the power of the court to entertain the suit.'" *U.S. v. Cambio Exacto, S.A.*, 166 F.3d 522, 526 (2d Cir. 1999) (quoting *In re Gucci*, 126 F.3d 380, 387-88 (2d Cir.1997)). Accordingly, I begin my inquiry here.

Plaintiffs argue that they have standing to pursue two forms of relief. *First*, Plaintiffs argue that they have standing to pursue damages because "the unlawful removal of CMI from a copyrighted work is a concrete injury." Opp. at 7. *Second*, Plaintiffs argue that they have standing to pursue injunctive relief, because they have alleged that there is a substantial risk that Defendants' program will "provide responses to users that incorporate[] material from Plaintiffs' copyright-protected works or regurgitate[] copyright-protected works verbatim or nearly verbatim." Compl. ¶ 52; *see also* Opp., at 9-10. Defendants respond that neither theory of harm identifies a concrete injury-in-fact sufficient to establish standing.

I agree with Defendants. Plaintiffs' claims for both damages and injunctive relief are DISMISSED because Plaintiffs' lack Article III standing. Accordingly, I need not reach the alternative motion pursuant to Fed. R. Civ. P. 12(b)(6).

I. Plaintiffs lack standing to pursue their claim for damages

Let us consider Plaintiffs' contention in respect of their claim for damages: that the unauthorized removal of CMI from their copyrighted work gives rise to a concrete injury-in-fact, even though they do not allege that a copy of their work from which the CMI has been removed has been disseminated by ChatGPT to anyone in response to any specific query. *See* Opp., at 7. Plaintiffs contend that their injury bears a "close relationship" to the tort of copyright infringement, because "the protection against removing or altering CMI, 17 U.S.C. § 1202(b)(i), is analogous to the rights to reproduce the works and prepare derivative ones, 17 U.S.C. §§ 106(1), (2)" in that "both grant the copyright owner the sole prerogative to decide how future iterations of the work may differ from the version the owner published." Opp., at 6. Plaintiffs argue that this, in turn, "accords with the common law, which recognizes interference with property, without more, as a concrete injury." *Id.* at 7.

I am not convinced that injury for interference with property provides the necessary "close historical or common-law analogue" to Plaintiffs' alleged injury. For one thing, Plaintiffs are wrong that Section 1202 "grant[s] the copyright owner the sole prerogative to decide how future iterations of the work may differ from the version the owner published." Other provisions of the Copyright Act afford such protections, *see* 17 U.S.C. § 106, but not Section 1202. Section 1202 protects copyright owners from specified interferences with the *integrity of a work's CMI*. In other words, Defendants may, absent permission, reproduce or even create derivatives of Plaintiffs' works—*without incurring liability under Section 1202*—as long as Defendants keep Plaintiffs' CMI intact. Indeed, the legislative history of the DMCA indicates that the Act's purpose was not to guard against property-based injury. Rather, it was to "ensure the integrity of the electronic marketplace by preventing fraud and misinformation," and to bring the United States into

compliance with its obligations to do so under the World Intellectual Property Organization (WIPO) Copyright Treaty, art. 12(1) ("Obligations concerning Rights Management Information") and WIPO Performances and Phonograms Treaty, art. 19 (same). H.R. Rep. No. 105-551, at 10-11.

Moreover, I am not convinced that the mere removal of identifying information from a copyrighted work—absent dissemination—has *any* historical or common-law analogue.

TransUnion is clear: "the plaintiff's injury [must] in fact be concrete—that is, real and not abstract. 594 U.S. at 424 (internal quotations and citations omitted). Plaintiffs allege that their copyrighted works (absent CMI) were used to train an AI-software program and remain in ChatGPT's repository of text. But Plaintiffs have not alleged any *actual* adverse effects stemming from this alleged DMCA violation. The argument advanced by Plaintiffs is akin to that of the *dissent* in *TransUnion*: "If a [defendant] breaches a [DMCA] duty owed to a specific [copyright owner], then that [copyright owner]…[has] a sufficient injury to sue in federal court." *Id.* at 450 (Thomas, J., dissenting). To this, the majority of the Court said: 'no.' "No concrete harm, no standing." *Id.* at 442. Accordingly, Plaintiffs lack Article III standing to seek retrospective relief in the form of damages for the injury they allege.

A. Plaintiffs lack standing to pursue their claim for injunctive relief

And so we turn to Plaintiffs' claim for injunctive relief. Plaintiffs seek an injunction "requiring Defendants to remove all copies of Plaintiffs' copyrighted works from which author, title, copyright, and terms of use information w[ere] removed from their training sets and any other repositories." Prayer for Relief.

Plaintiffs argue that they are entitled to such an injunction because, whether ChatGPT has or has not already reproduced their copyrighted work without attaching the required CMI, there is a substantial risk that ChatGPT will do so in the future. "[A] person exposed to a risk of future harm may pursue forward-looking, injunctive relief to prevent the harm from occurring, at least so long as the risk of harm is sufficiently imminent and substantial." *TransUnion*, 594 U.S. at 435. "An allegation of future injury may suffice if the threatened injury is 'certainly impending,' or there is a 'substantial risk that the harm will occur.'" *Susan B. Anthony List v. Driehaus*, 573 U.S. 149, 158 (2014) (quoting *Clapper v. Amnesty Int'l USA*, 568 U.S. 398, 414 n.5 (2013)). A substantial risk means there is a "realistic danger of sustaining a direct injury." *Pennell v. City of San Jose*, 485 U.S. 1, 8 (1988) (quoting *Babbitt v. United Farm Workers Nat'l Union*, 442 U.S. 289, 298 (1979)).

Defendants concede that there are clear historical and common-law analogues for this type of injury. Dkt. No. 71 ("Reply"), at 5 (citing *Masson v. New Yorker Mag., Inc.*, 501 U.S. 496, 511-12 (1991) (defamation); *Petrone v. Turner Publ'n Co. LLC*, 2023 WL 7302447, at *4 (S.D.N.Y. Nov. 6, 2023) (false light); *Gilliam v. Am. Broad. Cos., Inc.*, 538 F.2d 14, 24 (2d Cir. 1976) (unfair competition); *Lamothe v. Atl. Recording Corp.*, 847 F.2d 1403, 1406-07 (9th Cir. 1988) ("reverse passing off"); *Macia v. Microsoft Corp.*, 152 F. Supp. 2d 535, 541 (D. Vt. 2001) (slander of title)). However, Defendants argue that Plaintiffs lack standing to seek injunctive relief because they fail to allege facts tending to show that the risk of ChatGPT reproducing Plaintiffs' work, in whole or in part, absent the requisite CMI is "substantial." *See* Reply, at 3.

I agree with Defendants. Plaintiffs allege that ChatGPT has been trained on "a scrape of most of the internet," Compl. ¶ 29, which includes massive amounts of information from innumerable sources on almost any given subject. Plaintiffs have nowhere alleged that the

information in their articles is copyrighted, nor could they do so. When a user inputs a question into ChatGPT, ChatGPT synthesizes the relevant information in its repository into an answer. Given the quantity of information contained in the repository, the likelihood that ChatGPT would output plagiarized content from one of Plaintiffs' articles seems remote. And while Plaintiffs provide third-party statistics indicating that an *earlier* version of ChatGPT generated responses containing significant amounts of plagiarized content, Compl. ¶ 5, Plaintiffs have not plausibly alleged that there is a "substantial risk" that the *current* version of ChatGPT will generate a response plagiarizing one of *Plaintiffs'* articles.

Accordingly, Plaintiffs lack Article III standing to seek injunctive relief for their alleged injury.

* * *

Let us be clear about what is really at stake here. The alleged injury for which Plaintiffs truly seek redress is not the exclusion of CMI from Defendants' training sets, but rather Defendants' *use* of Plaintiffs' articles to develop ChatGPT without compensation to Plaintiffs. *See* Compl. ¶ 57 ("The OpenAI Defendants have acknowledged that use of copyright-protected works to train ChatGPT requires a license to that content, and in some instances, have entered licensing agreements with large copyright owners…They are also in licensing talks with other copyright owners in the news industry, but have offered no compensation to Plaintiffs."). Whether or not that type of injury satisfies the injury-in-fact requirement, it is not the type of harm that has been "elevated" by Section 1202(b)(i) of the DMCA. *See Spokeo,* 578 U.S. at 341 (Congress may "elevate to the status of legally cognizable injuries, *de facto* injuries that were previously inadequate in law."). Whether there is another statute or legal theory that does elevate this type of harm remains to be seen. But that question is not before the Court today.

In the event of dismissal Plaintiffs seek leave to file an amended complaint. I cannot ascertain whether amendment would be futile without seeing a proposed amended pleading. I am skeptical about Plaintiffs' ability to allege a cognizable injury but, at least as to injunctive relief, I am prepared to consider an amended pleading.

Conclusion

For the foregoing reasons, Defendants' motion to dismiss is GRANTED in its entirety. Plaintiffs' motion for leave to replead is DENIED WITHOUT PREJUDICE to renewal on a proper record—which means filing a notice of motion to which a proposed amended pleading is attached together with an explanation of why the proposed amendment would not be futile.

This constitutes a written opinion. The Clerk is directed to remove the motion at Dkt. No. 68 from the Court's list of open motions.

Dated: November 7, 2024

U.S.D.J.

BY ECF TO ALL COUNSEL

EXHIBIT 2

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, OPENAI HOLDINGS, LLC,

Defendants.

Civil Action No. 24-01514-CM

FIRST AMENDED COMPLAINT

JURY TRIAL DEMANDED

1. Plaintiffs Raw Story Media, Inc. and AlterNet Media, Inc., through their attorneys Loevy & Loevy, for their Complaint against the OpenAI Defendants, allege the following:

2. The Copyright Clause of the U.S. Constitution empowers Congress to protect works of human creativity. The resulting legal protections encourage people to devote effort and resources to all manner of creative enterprises by providing confidence that creators' works will be shielded from unauthorized encroachment.

3. In recognition that emerging technologies could be used to evade statutory protections, Congress passed the Digital Millennium Copyright Act in 1998. The DMCA prohibits the removal of author, title, copyright, and terms of use information from protected works where there is reason to know that it would induce, enable, facilitate, or conceal a copyright infringement. Unlike copyright infringement claims, which require copyright owners to incur significant and often prohibitive registration costs as a prerequisite to enforcing their copyrights, a DMCA claim does not require registration.

4. Generative artificial intelligence (AI) systems and large language models (LLMs) are trained using works created by humans. AI systems and LLMs ingest massive amounts of human creativity and use it to mimic how humans write and speak. These training sets have included hundreds of thousands, if not millions, of works of journalism.

5. Defendants are the companies primarily responsible for the creation and development of the highly lucrative ChatGPT AI products. According to the award-winning website Copyleaks, nearly 60% of the responses provided by Defendants' GPT-3.5 product in a study conducted by Copyleaks contained some form of plagiarized content, and over 45% contained text that was identical to pre-existing content.

6. When they populated their training sets with works of journalism, Defendants had a choice: they could train ChatGPT using works of journalism with the copyright management information protected by the DMCA intact, or they could strip it away. Defendants chose the latter, and in the process, trained ChatGPT not to acknowledge or respect copyright, not to notify ChatGPT users when the responses they received were protected by journalists' copyrights, and not to provide attribution when using the works of human journalists.

7. Plaintiffs Raw Story and AlterNet are news organizations, and bring this lawsuit seeking actual damages and Defendants' profits, or statutory damages of no less than $2500 per violation.

PARTIES

8. For over two decades, Raw Story has published award-winning investigative journalism, breaking news, and bold opinion columns. Raw Story publishes to more than ten million readers each month and has more than 1,000,000 daily readers. It is the largest independent

progressive political news website in America and was named the best news/political blog in America by *Editor & Publisher* in 2022 and 2023.

9. Among other important work, Raw Story has received *Editor & Publisher* (EPPY), Society of Professional Journalists, Fair Media Council, and ION awards for its reporting on white nationalism, the January 6 riots, South Dakota governor Kristi Noem's use of a state airplane for non-official purposes, and inappropriate Congressional stock trading. Raw Story reporters produce timely, illuminating work at great risk and cost to enrich understanding of critical issues and undermine threats to civil society.

10. As one example of the risks taken by Raw Story reporters in bringing the news to the public—risks never faced by AI bots—members of a neo-Nazi group showed up at the home of a Raw Story reporter who covers extremism and white supremacy in America. *See* Washington Post, *A reporter investigated neo-Nazis. Then they came to his house in masks.* (Feb. 20, 2024), https://www.washingtonpost.com/style/media/2024/02/20/raw-story-neo-nazi-journalist-house/.

11. Raw Story is a Massachusetts corporation with its headquarters in Miami Beach, Florida.

12. AlterNet is a three-time Webby award-winning publisher with a focus on civil rights, social justice, culture, health, and the environment. For 25 years, AlterNet's reporters have chased political news, and its opinion writers have probed the intersection of politics, science, and religion.

13. AlterNet is a Delaware corporation with its headquarters in Miami Beach, Florida.

14. Together, Plaintiffs have published more than 400,000 breaking news features, investigative news articles and opinion columns as a result of their considerable investments of time and resources.

15. Defendants are the inter-related organizations primarily responsible for the creation, training, marketing, and sale of ChatGPT AI products.

16. OpenAI Inc. is a Delaware nonprofit corporation with a principal place of business in San Francisco, CA.

17. OpenAI OpCo LLC is a Delaware limited liability company with a principal place of business in San Francisco, CA. OpenAI OpCo LLC is the sole member of OpenAI, LLC. Previously, OpenAI OpCo was known as OpenAI LP.

18. OpenAI GP, LLC is a Delaware limited liability company with a principal place of business in San Francisco, CA. It is the general partner of OpenAI OpCo and controls OpenAI OpCo.

19. OpenAI, LLC is a Delaware limited liability company with a principal place of business in San Francisco, CA. It owns some of the services or products operated by OpenAI.

20. OpenAI Global LLC is a Delaware limited liability company with a principal place of business in San Francisco, CA.

21. OAI Corporation, LLC is a Delaware limited liability company with a principal place of business in San Francisco, CA. Its sole member is OpenAI Holdings, LLC.

22. OpenAI Holdings, LLC is a Delaware limited liability company with a principal place of business in San Francisco, CA. Its sole members are OpenAI, Inc. and Aestas Corporation.

JURISDICTION AND VENUE

23. The Court has subject matter jurisdiction under 28 U.S.C. §§ 1331 and 1338(a) because this action arises under the Copyright Act of 1976, 17 U.S.C. § 101, et seq., as amended by the Digital Millennium Copyright Act.

24. Jurisdiction over Defendants is proper because they have purposefully availed themselves of New York to conduct their business. OpenAI maintains offices and employs staff in New York who, on information and belief, were engaged in training and/or marketing OpenAI's GenAI systems and LLMs, and thus in the removal of Plaintiffs' copyright management information as discussed in this Complaint and/or the sale of products to New York residents resulting from that removal. Defendants did not contest personal jurisdiction in their Motion to Dismiss.

25. Venue is proper under 28 U.S.C. § 1400(a) because Defendants or their agents reside or may be found in this District.

26. Venue is also proper under 28 U.S.C. § 1391(b)(2) because a substantial part of the acts or omissions giving rise to Plaintiffs' claims occurred in this District. Specifically, OpenAI employs staff in New York who, on information and belief, were engaged in the activities alleged in this Complaint.

27. Defendants did not contest venue in their Motion to Dismiss.

PLAINTIFFS' COPYRIGHT-PROTECTED WORKS OF JOURNALISM

28. Plaintiffs' copyrighted works of journalism are published on Plaintiffs' websites, rawstory.com and alternet.org respectively, and are conveyed to the public with author, title, and copyright notice information.

29. Plaintiffs own copyrights to all the Raw Story articles listed in Exhibit 1 and the AlterNet articles listed in Exhibit 2.

30. Plaintiffs' copyright-protected works are the result of significant investment by Plaintiffs in the human and other resources necessary to report on the news.

DEFENDANTS' INCLUSION OF PLAINTIFFS' WORKS IN THEIR TRAINING SETS AND REMOVAL OF COPYRIGHT MANAGEMENT INFORMATION

31. Defendants' generative AI products utilize a "large language model," or "LLM." The different versions of GPT are examples of LLMs. An LLM, including those that power ChatGPT, take text prompts as inputs and emit outputs to predict responses that are likely to follow a given the potentially billions of input examples used to train it.

32. LLMs arrive at their outputs as the result of their training on works written by humans, which are often protected by copyright. They collect these examples in training sets.

33. When assembling training sets, LLM creators, including Defendants, first identify the works they want to include. They then encode the work in computer memory as numbers called "parameters."

34. Defendants have not published the contents of the training sets used to train any version of ChatGPT, but have disclosed information about those training sets prior to GPT-4.[1] Beginning with GPT-4, Defendants have been fully secret about the training sets used to train that and later versions of ChatGPT. Plaintiffs' allegations about Defendants' training sets are therefore based upon an extensive review of publicly available information regarding earlier versions of ChatGPT and consultations with a data scientist employed by Plaintiffs' counsel to analyze that information and provide insights into the manner in which AI is developed and functions.

35. Plaintiffs' allegations about Defendants' training sets are also based upon information learned through discovery in this case.

36. Earlier versions of ChatGPT (prior to GPT-4) were trained using at least the following training sets: WebText, WebText2, and sets derived from Common Crawl.

[1] Plaintiffs collectively refer to all versions of ChatGPT as "ChatGPT" unless a specific version is specified.

37. WebText and WebText2 were created by Defendants. They are collections of all outbound links on the website Reddit that received at least three "karma."[2] On Reddit, a karma indicates that users have generally approved the link. The difference between the datasets is that WebText2 involved scraping links from Reddit over a longer period of time. Thus, WebText2 is an expanded version of WebText.

38. Defendants have published a list of the top 1,000 web domains present in the WebText training set and their frequency. According to that list, WebText included 33,598 distinct URLs from Raw Story's web domain and 23,183 distinct URLs from AlterNet's web domain.[3]

39. Defendants have a record of, and are aware, of each URL that was included in each of their training sets.

40. Joshua C. Peterson, currently an assistant professor in the Faculty of Computing and Data Sciences at Boston University, and two computational cognitive scientists with PhDs from U.C. Berkeley, created an approximation of the WebText dataset, called OpenWebText, by also scraping outbound links from Reddit that received at least three "karma," just like Defendants did in creating WebText.[4] They published the results online. A data scientist employed by Plaintiffs' counsel then analyzed those results. OpenWebText contains 40,403 distinct URLs from rawstory.com and 34,003 from alternet.org. A list of the Raw Story works contained in OpenWebText is attached as Exhibit 3. A list of the AlterNet works contained in OpenWebText is attached as Exhibit 4.

---

[2] Alec Radford et al, Language Models are Unsupervised Multitask Learners, 3, https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[3] https://github.com/openai/gpt-2/blob/master/domains.txt.

[4] https://github.com/jcpeterson/openwebtext/blob/master/README.md.

41. Upon information and belief, there are different numbers of Plaintiffs' articles in WebText and OpenWebText at least in part because the scrapes occurred on different dates.

42. Defendants have explained that, in developing WebText, they used sets of algorithms called Dragnet and Newspaper to extract text from websites.[5] Upon information and belief, Defendants used these two extraction methods, rather than one method, to create redundancies in case one method experienced a bug or did not work properly in a given case. Applying two methods rather than one would lead to a training set that is more consistent in the kind of content it contains, which is desirable from a training perspective.

43. Dragnet's algorithms are designed to "separate the main article content" from other parts of the website, including "footers" and "copyright notices," and allow the extractor to make further copies only of the "main article content."[6]

44. Like Dragnet, the Newspaper algorithms are incapable of extracting copyright notices and footers. Further, a user of Newspaper has the choice to extract or not extract author and title information. On information and belief, Defendants chose not to extract author and title information because they desired consistency with the Dragnet extractions, and Dragnet is designed not to extract author and title information.

45. In applying the Dragnet and Newspaper algorithms while assembling the WebText dataset, Defendants removed Plaintiffs' author, title, and copyright notice information.

---

[5] Alec Radford et al., Language Models are Unsupervised Multitask Learners, 3 https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

[6] Matt McDonnell, Benchmarking Python Content Extraction Algorithms (Jan. 29, 2015), https://moz.com/devblog/benchmarking-python-content-extraction-algorithms-dragnet-readability-goose-and-eatiht.

46. Upon information and belief, Defendants, when using Dragnet and Newspaper, first download and save the relevant webpage before extracting data from it. This is at least because, when they use Dragnet and Newspaper, they likely anticipate a possible future need to regenerate the dataset (*e.g.*, if the dataset becomes corrupted), and it is cheaper to save a copy than it is to recrawl all the data.

47. Because, by the time of its scraping, Dragnet and Newspaper were publicly known to remove author, title, copyright notices, and footers, and given that OpenAI employs highly skilled data scientists who would know how Dragnet and Newspaper work, Defendants intentionally and knowingly removed this copyright management information while assembling WebText.

48. A data scientist employed by Plaintiffs' counsel applied the Dragnet code to Raw Story and AlterNet URLs. As a representative sample, three results from Raw Story and three results from Alternet are attached as Exhibit 5. Of the examples Plaintiffs have examined, the resulting copies' text is in some cases completely identical to the original and in other substantively identical to the original (e.g., completely identical except for the seemingly random addition of an extra space between two words, the omission of a description or credit associated with an embedded photo, or the alteration of some formatting). Of the examples Plaintiffs have examined, every copy lacks the copyright notice information with which it was originally conveyed, while most copies lack author and title.

49. A data scientist employed by Plaintiffs' counsel also applied the Newspaper code to Raw Story and AlterNet URLs contained in OpenWebText. The data scientist applied the version of the code that enables the user not to extract author and title information based on the reasonable assumption that Defendants desired consistency with the Dragnet extractions. As a

representative sample, three results for Raw Story and three results for AlterNet are attached as Exhibit 6. Of the examples Plaintiffs have examined, the resulting copies' text is in some cases completely identical to the original and in others substantively identical to the original in the same sense as the Dragnet extractions (except that Newspaper does not randomly add spaces), while in one case a bulleted list was removed. Of the examples Plaintiffs have examined, every copy lacks the author, title, and copyright notice information with which it was conveyed to the public.

50. Plaintiffs' data scientist did not alter the article in any way other than by applying the Dragnet or Newspaper algorithm. Thus, both the removal of CMI and any trivial alterations to the article occurred simultaneously by applying the Dragnet or Newspaper algorithm to an identical copy of Plaintiffs' work.

51. The absence of author, title, and copyright notice information from the copies of Plaintiffs' articles generated by applying the Dragnet and Newspaper codes—codes OpenAI has admitted to have intentionally used when assembling WebText—further corroborates that Defendants intentionally removed author, title, and copyright notice information from Plaintiffs' copyright-protected news articles.

52. Upon information and belief, Defendants have continued to use the same or similar Dragnet and Newspaper text extraction methods when creating training sets for every version of ChatGPT since GPT-2. This is at least because Defendants have admitted to using these methods for GPT-2 and have neither publicly disclaimed their use for later version of ChatGPT nor publicly claimed to have used any other text extraction methods for those later versions.

53. Common Crawl is a data set that consists of a scrape of most of the internet created by a non-profit research institute, also called Common Crawl. ChatGPT was trained on a version of Common Crawl, in addition to the WebText and WebText2 training sets.

54. To train GPT-2, OpenAI downloaded Common Crawl data from the third party's website and filtered it to include only certain works, such as those written in English.[7] When the third party downloads the article into Common Crawl, author, title, and copyright notice information is preserved.

55. Google has published instructions on how to replicate a dataset called C4, a monthly snapshot of filtered Common Crawl data that Google used to train its own AI models. Upon information and belief, based on the similarity of Defendants' and Google's goals in training AI models, C4 is substantially similar to the filtered versions of Common Crawl used to train ChatGPT. The Allen Institute for AI, a nonprofit research institute launched by Microsoft cofounder Paul Allen, followed Google's instructions and published its recreation of C4 online.[8]

56. A data scientist employed by Plaintiffs' counsel analyzed this recreation. It contains 8,974 distinct URLs from rawstory.com. The vast majority of these URLs contain Plaintiffs' copyright-protected news articles. None of the news articles contains copyright notice information. Most lack both author and title information. In some cases, the articles' text is completely identical to the original. In other cases, the articles' text is substantively identical to the original in the same sense as the Dragnet extractions (except that the C4 algorithm does not randomly add spaces), while in still others a small number of paragraphs are omitted.

[7] Tom B. Brown et al, Language Models are Few-Shot Learners, 14 (July 22, 2020), https://arxiv.org/pdf/2005.14165.
[8] https://huggingface.co/datasets/allenai/c4.

57. Upon information and belief, both the removal of CMI and any trivial alterations to the article occurred simultaneously by applying the C4 algorithm to an identical copy of Plaintiffs' work.

58. As a representative sample, the text of three Raw Story articles as they appear in the C4 set is attached as Exhibit 7. None of these articles contains the author, title, or copyright notice information with which it was conveyed to the public.

59. In discovery, Defendants produced documents showing that Raw Story and AlterNet articles were included in Defendants' training sets. Tables listing each article's URL and the number of total occurrences across the training sets are attached as Exhibit 8 and Exhibit 9 for each article that Plaintiffs own.

60. Plaintiffs have examined a sample of the copies as they appear in the training sets. None of the copies in the examined sample contains the author, title, or copyright notice with which the original was conveyed. Apart from the removal of author, title, and copyright notice information, and the inclusion of certain symbols for formatting purposes (for example, the use of "/n/n" to denote a new paragraph), the text of the training set copies is identical to that of the originals in the sample examined by Plaintiffs.

61. Plaintiffs have not licensed or otherwise permitted Defendants to include any of their works in their training sets.

62. Defendants' actions in downloading thousands of Plaintiffs' articles without permission infringes Plaintiffs' copyright, more specifically, the right to control reproductions of copyright-protected works.

DEFENDANTS' REGURGITATION, MIMICKING, AND ABRIDGEMENT OF COPYRIGHT-PROTECTED WORKS OF JOURNALISM

63. ChatGPT offers a product to its customers that provides responses to questions or other prompts. ChatGPT's ability to provide these responses is the key value proposition of its product, one which it is able to sell to its customers for enormous sums of money, soon likely to be in the billions of dollars.

64. To train ChatGPT, Defendants retain users' chat histories with ChatGPT unless the user takes the affirmative step of disabling that feature.[9] Thus, upon information and belief, Defendants possess a repository of every regurgitation or abridgement of Plaintiffs' works apart from those whose storage users have affirmatively disabled.

65. At least some of the time, ChatGPT provides or has provided responses to users that regurgitate verbatim or nearly verbatim copyright-protected works of journalism without providing any author, title, or copyright notice information conveyed in connection with those works. Examples of such regurgitations are included in Exhibit J to the Complaint in *Daily News, LP v. Microsoft Corporation*, No. 24-cv-03285 (S.D.N.Y. Apr. 30, 2024), ECF No. 1.

66. At least some of the time, ChatGPT provides or has provided responses to users that mimic significant amounts of material from copyright-protected works of journalism without providing any author, title, or copyright notice information contained in those works. For example, if a user asks ChatGPT about a current event or the results of a work of investigative journalism, ChatGPT will provide responses that mimic copyright-protected works of journalism that covered those events, not responses that are based on any journalism efforts by Defendants.

[9] New ways to manage your data in ChatGPT (Apr. 25, 2023), https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt/.

67. At least some of the time, ChatGPT memorizes and regurgitates material.[10] Defendants have publicly admitted their knowledge of this fact. Defendants have also effectively admitted that regurgitation of copyrighted works is infringement: when Plaintiffs attempted to obtain the same regurgitations set forth in the *Daily News* case using the same methodology, Plaintiffs received in one instance a message stating, "I'm sorry, but I can't generate the original ending for the article or any copyrighted content." Thus, upon information and belief, Defendants have recently changed ChatGPT to reduce regurgitations for copyright reasons.

68. At least some of the time, ChatGPT provides or has provided responses to users that abridge copyright-protected works of journalism without providing any author, title, or copyright notice information conveyed in connection with those works. Examples of such abridgements are included in Exhibit 11 to the First Amended Complaint in *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-cv-4872 (S.D.N.Y. Sept. 9, 2024), ECF No. 88-14. For instance, in the fourth example, the ChatGPT abridgement reproduces, verbatim, nine consecutive paragraphs of text (minus one sentence) from the original article, which can be found at https://www.motherjones.com/politics/2024/01/100-bill-crime-corruption-treasury-tax-evasion/.

69. When earlier versions of ChatGPT (up to and including ChatGPT 3.5-turbo) abridge a copyright-protected news article in response to a user prompt, they draw from their training on the article. During training, the patterns of all content, including copyright-protected news articles, are imprinted onto the model. That imprint allows the model to abridge the article.

70. When later versions of ChatGPT abridge a copyright-protected news article in response to a user prompt, they find the previously downloaded article inside a database called a

[10] OpenAI and journalism (Jan. 8, 2024), https://openai.com/index/openai-and-journalism/.

search index using a method called synthetic searching or retrieval-augmented generation ("RAG"). Upon information and belief, they make another copy of the article in the memory of their computing system and use their LLM or other programming to generate an abridgement by applying the model or other programming to the text of the article.

71. Plaintiffs' articles are not merely collections of facts. Rather, they reflect the originality of their authors in selecting, arranging, and presenting facts to tell compelling stories. They also reflect the authors' analysis and interpretation of events, structuring of materials, marshaling of facts, and the emphasis given to certain aspects.

72. An ordinary observer of a ChatGPT abridgement of copyright-protected news articles would conclude that the abridgements were derived from the articles being abridged.

73. In response to prompts seeking an abridgement of an article, ChatGPT will typically provide a general abridgement of such an article, on the order of a few paragraphs. In some instances, the initial response will summarize the article in substantial detail. Further, when prompted by the user to provide more information about one or more aspects of that abridgement, ChatGPT will provide additional details, often in the format of a bulleted list of main points. If prompted by the user to provide more information on one of more of those points, ChatGPT will provide additional details. In some instances, however, ChatGPT will provide a bulleted list of main points in response to an initial prompt seeking an abridgement.

74. A ChatGPT user is capable of obtaining a substantial abridgement of a copyright protected news article through such series of prompts, and in some instances, further prompts designed to elicit further summary are even suggested by ChatGPT itself. *See* Exhibit 11 to the First Amended Complaint in *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-cv-4872 (S.D.N.Y. Sept. 9, 2024), ECF No. 88-14. These abridgements lack copyright notice

information conveyed in connection with the work, and sometimes lack author information. They sometimes link to webpages that do not belong to the news organization that owns the article and that do not contain the news organization's copyright management information.

75. Thus, upon information and belief, abridgements from earlier versions of ChatGPT lack copyright notice and typically author information because Defendants intentionally removed that information from the ChatGPT training sets.

76. Further, the abridgements from later versions of ChatGPT lack copyright notice and typically author information. Upon information and belief, this is because Defendants intentionally removed them either when initially storing them in computer memory or when generating results by employing RAG.

77. On April 22, 2024, Plaintiffs served a Request for Production on Defendants seeking all chat responses that included any portion of any articles published on rawstory.com or alternet.org. That same day, Plaintiffs also served a Request for Production on Defendants seeking all chat responses that regurgitated any articles from rawstory.com or alternet.org.

78. Defendants agreed to produce this information many months ago, but despite Plaintiffs' diligent follow-ups, they have not yet done so. Analysis of this information would show the extent to which Defendants provided regurgitations or other outputs that sufficiently disseminated Plaintiffs' works or abridgement of those works.

DEFENDANTS' INTENTIONAL REMOVAL OF COPYRIGHT MANAGEMENT INFORMATION FROM PLAINTIFFS' WORKS IN THEIR TRAINING SETS

79. ChatGPT does not have any independent knowledge of the information provided in its responses. Rather, to service Defendants' paying customers, ChatGPT instead repackages, among other material, the copyrighted journalism work product developed by Plaintiffs and others at their expense.

80. When providing responses, ChatGPT gives the impression that it is an all-knowing, "intelligent" source of the information being provided, when in reality, the responses are frequently based on copyrighted works of journalism that ChatGPT simply mimics.

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 when providing responses to users unless Defendants trained it otherwise.

82. Based on the publicly available information described above, thousands of Plaintiffs' copyrighted works were included in Defendants' training sets without the author, title, and copyright notice information that Plaintiffs conveyed in publishing them.

83. Based on the publicly available information described above, including Defendants' admission to using the Dragnet and Newspaper extraction methods, which remove author, title, and copyright notice information from copyright-protected news articles published online, Defendants intentionally removed author, title, and copyright notice information from Plaintiffs' copyrighted works in creating ChatGPT training sets.

DEFENDANTS' ACTUAL AND CONSTRUCTIVE KNOWLEDGE OF THEIR VIOLATIONS

84. Defendants have acknowledged that use of copyright-protected works to train ChatGPT requires a license to that content. and, in some instances. Recognizing that obligation, Defendants have entered into agreements with large copyright owners such as Associated Press, the Atlantic, Axel Springer, Dotdash Meredith, Financial Times, News Corp, and Vox Media to obtain licenses to include those entities' copyright-protected works in Defendants' LLM training data.

85. Defendants are also in licensing talks with other copyright owners in the news industry, but have offered no compensation to Plaintiffs.

86. In a May 29, 2024 interview, OpenAI's Chief of Intellectual Property and Content, Tom Rubin, stated that these deals focus on "the display of news content and use of the tools and tech," and are thus "largely not" about training.[11] This admission, while qualified, confirms that these deals involve training, at least in part.

87. Defendants created tools in late 2023 to allow copyright owners to block their work from being incorporated into training sets. This further corroborates that Defendants had reason to know that use of copyrighted material in their training sets without permission or license is copyright infringement.

88. The creation of such tools also corroborates that Defendants had reason to know that their copyright infringement is enabled, facilitated, and concealed by their removal of author, title, and copyright notice information from their training sets.

89. Defendants had reasonable grounds to know that the removal of author, title, and copyright notice information from copyright-protected works and their use in training ChatGPT would result in ChatGPT providing responses to ChatGPT users that incorporated or regurgitated material verbatim from copyrighted works in creating responses to users, without revealing that those works were subject to Plaintiffs' copyrights. This is at least because Defendants were aware that ChatGPT responses are the product of its training sets and that ChatGPT would not know any author, title, and copyright information that was not included in training sets.

90. Upon information and belief, Defendants had reason to know that the removal of author, title, and copyright notice information from copyright-protected works used in synthetic searching would result in ChatGPT providing responses to ChatGPT users that abridged or

[11] Charlotte Tobitt, OpenAI content boss: 'Incumbent' on us to help small publishers, not just the giants, *PressGazette* (May 30, 2024), https://pressgazette.co.uk/platforms/openai-tom-rubin-publishers-news/.

regurgitated material verbatim from copyrighted works in creating responses to users, without revealing that those works were subject to Plaintiffs' copyrights. This is at least because Defendants were aware that later versions of ChatGPT's responses to prompts are the product of the articles encoded in their computer memory, from which, upon information and belief, Defendants removed author, title, and copyright notice information.

91. Defendants had reason to know that users of ChatGPT would further distribute the results of ChatGPT responses. This is at least because Defendants promote ChatGPT as a tool that can be used by a user to generate content for a further audience.

92. Defendants had reason to know that users of ChatGPT would be less likely to distribute ChatGPT responses if they were made aware of the author, title, and copyright notice information applicable to the material used to generate those responses. This is at least because Defendants were aware that at least some likely users of ChatGPT respect the copyrights of others or fear liability for copyright infringement.

93. Defendants had reason to know that ChatGPT would be less popular and would generate less revenue if users believed that ChatGPT responses violated third-party copyrights or if users were otherwise concerned about further distributing ChatGPT responses. This is at least because Defendants were aware that they derive revenue from user subscriptions, that at least some likely users of ChatGPT respect the copyrights of others or fear liability for copyright infringement, and that such users would not pay to use a product that might result in copyright liability or did not respect the copyrights of others.

94. If a commercial user of ChatGPT is sued for copyright infringement, Defendants have committed to paying the user's costs in defending against the infringement claim, and to indemnifying the user for an adverse judgment or settlement. These commitments apply only if

the user uses the product as advertised. In particular, OpenAI's "Copyright Shield" does not apply if the user "disabled, ignored, or did not use any relevant citation, filtering or safety features or restrictions provided by OpenAI."[12] Thus, Defendants know or have reason to know that ChatGPT users are capable of infringing and likely to infringe copyright even when used according to terms specified by Defendants.

95. Defendants intend in part for ChatGPT to replicate how ordinary English speakers express themselves. When ordinary English speakers are not conveying copyright-protected works, they do not include copyright management information—especially copyright notices. Had ChatGPT been trained on Plaintiffs' and others' copyright-protected works that include this copyright management information, it would have falsely learned that ordinary English speakers convey copyright management information in situations when they do not. To avoid this result, Defendants had a choice between removing the copyright management information at the outset or retraining ChatGPT not to emit the copyright management information after it had incorrectly learned how English speakers normally express themselves. Upon information and belief, Defendants chose to remove the copyright management information at the outset, at least because doing so involves fewer computational resources and therefore is far less expensive than retraining. Thus, because Defendants infringed Plaintiffs' copyright by using Plaintiffs' articles to train ChatGPT, Defendants removed Plaintiffs' copyright management information from its copyright-protected articles knowing, or having reasonable grounds to know, that doing so would facilitate their own training-based infringing conduct.

[12] https://openai.com/policies/service-terms/.

96. Defendants' unauthorized copying of Plaintiffs' works into Defendants' training sets and search indices is facilitated by the removal of author, title, and copyright notice information because copying less data requires fewer computational and storage resources.

DEFENDANTS' CONTINUING VIOLATIONS

97. Upon information and belief, Defendants have continued to unlawfully copy, regurgitate, abridge, and remove author, title, and copyright notice information from Plaintiffs' copyright-protected works up to the present date, or at least until Plaintiffs implemented the exclusion protocols on October 27, 2023, that Defendants released in August 2023 allowing websites to opt out of OpenAI's web crawling.

98. ChatGPT has emitted significant material from copyright-protected works of journalism that significantly postdate the WebText and WebText2 training sets. Examples are contained in Exhibit 11 to the First Amended Complaint in *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 24-cv-4872 (S.D.N.Y. Sept. 9, 2024), ECF No. 88-14. ChatGPT could not have produced this material without Defendants' copying the original articles and storing them in computer memory, including in training sets created by ChatGPT 3.5-turbo and earlier, and search indices for RAG purposes.

99. In addition, each successive GPT model has had orders of magnitude more parameters than the last. For instance, GPT-4 is reported to have 1.8 trillion parameters,[13] a tenfold increase from the 175 billion parameters used to train GPT-3.[14] Because adding more parameters requires training on more data, it is unlikely that Defendants would have foregone including

[13] Maximilian Schreiner, GPT-4 architecture, datasets, costs and more leaked, The Decoder (July 11, 2023), https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/.
[14] Tom B. Brown et al, Language Models are Few-Shot Learners, 5 (July 22, 2020), https://arxiv.org/pdf/2005.14165.

Plaintiffs' articles in their more recent training sets. Thus, upon information and belief, Defendants continue to include Plaintiffs' articles in their training sets up to the present date.

100. Further, Defendants' adoption of a tool in August 2023 to allow website owners to block web crawling would have been unnecessary they were not continuing to copy content from the internet, including Plaintiffs' copyright-protected works, as they had done in the past.

101. According to OpenAI's Chief of Intellectual Property and Content, each of OpenAI's models is "trained from scratch."[15] Thus, when assembling new training sets, OpenAI recrawls the same articles it included in past training sets.

102. As alleged above, upon information and belief, Defendants have continued to use the same or similar Dragnet and Newspaper text extraction methods when creating training sets for every version of ChatGPT since GPT-2. Thus, upon information and belief, they have continued to remove author, title, and copyright notice information from Plaintiffs' copyright-protected articles up to the present, including but not limited to Plaintiffs' articles that are contained in Defendants' training sets created in the past three years.

Count I – Violation of 17 U.S.C. § 1202(b)(1)

103. The above paragraphs are incorporated by reference into this Count.

104. Plaintiffs are the owners of copyrighted works of journalism that contain author, title, and copyright notice information.

105. Defendants created copies of Plaintiffs' works of journalism with author information removed and included them in training sets used to train ChatGPT.

[15] Charlotte Tobitt, OpenAI content boss: 'Incumbent' on us to help small publishers, not just the giants, PressGazette (May 30, 2024), https://pressgazette.co.uk/platforms/openai-tom-rubin-publishers-news/.

106. Defendants created copies of Plaintiffs' works of journalism with title information removed and included them in training sets used to train ChatGPT.

107. Defendants created copies of Plaintiffs' works of journalism with copyright notice information removed and included them in training sets used to train ChatGPT.

108. Defendants had reason to know that inclusion in their training sets of Plaintiffs' works of journalism without author, title, and copyright notice information would induce ChatGPT to provide responses to users that incorporated material from Plaintiffs' copyright-protected works, and abridged or regurgitated copyright-protected works verbatim or nearly verbatim.

109. Defendants had reason to know that inclusion in their training sets of Plaintiffs' works of journalism without author, title, and copyright notice information would induce ChatGPT users to distribute or publish ChatGPT responses that utilized Plaintiffs' copyright-protected works of journalism that such users would not have distributed or published if they were aware of the author, title, and copyright notice information.

110. Defendants had reason to know that inclusion in their training sets of Plaintiffs' works of journalism without author, title, and copyright notice information would enable copyright infringement by Defendants, ChatGPT and ChatGPT users.

111. Defendants had reason to know that inclusion in their training sets of Plaintiffs' works of journalism without author, title, and copyright notice information would facilitate copyright infringement by Defendants, ChatGPT and ChatGPT users.

112. Defendants had reason to know that inclusion in their training sets of Plaintiffs' works of journalism without author, title, and copyright notice information would conceal copyright infringement by Defendants, ChatGPT, and ChatGPT users.

PRAYER FOR RELIEF

Plaintiffs seek the following relief:

(i) Either statutory damages or the total of Plaintiffs' damages and Defendants' profits, to be elected by Plaintiffs;

(ii) An injunction requiring Defendants to remove all copies of Plaintiffs' copyrighted works from which author, title, or copyright notice information was removed from their training sets and any other repositories;

(iii) An injunction prohibiting the unlawful conduct alleged above;

(iv) An injunction ordering the destruction of all GPT or other LLMs and training sets that incorporate Plaintiffs' works from which author, title, or copyright notice has been removed; and

(v) Attorney fees and costs.

JURY DEMAND

Plaintiffs demand a jury trial.

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

EXHIBIT 3

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

x

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

Plaintiffs,

vs.

24 Civ. 01514

OPENAI, INC., OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, and OPENAI HOLDINGS, LLC,

Defendants.

x

DECISION AND ORDER

McMahon, J.:

This is a continuation of the action brought by Plaintiffs Raw Story Media, Inc. and AlterNet Media, Inc. (collectively "Plaintiffs") against OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI Opco LLC, OpenAI Global LLC, and OpenAI Holdings, LLC (collectively, "Defendants" or "OpenAI"), pursuant to the Digital Millennium Copyright Act (the "DMCA"), 17 U.S.C. § 1201 *et seq*.

Plaintiffs Raw Story Media, Inc. and AlterNet Media, Inc. allege that "thousands" of their copyright-protected works of journalism were stripped of their copyright management information (CMI), input into at least three of OpenAI's training sets (WebText, WebText2, and Common Crawl), which were then used to train ChatGPT in violation of Section 1202(b)(i) of the DMCA. Dkt. No. 1, ¶¶ 14, 24, 29.

On November 7, 2024, I granted OpenAI's motion to dismiss Plaintiffs' complaint, pursuant to Fed. R. Civ. P. 12(b)(1), holding that Plaintiffs had not alleged an injury-in-fact sufficient to confer standing. Dkt. No. 117.

Plaintiffs have moved for leave to amend their complaint pursuant to Fed. R. Civ. P. 15(a)(2), or in the alternative, for leave to continue taking jurisdictional discovery.

For the reasons below, Plaintiffs' motion is DENIED.[1]

* * *

"No concrete harm, no standing." *TransUnion LLC v. Ramirez*. 594 U.S. 413, 417, 442 (2021). This the Supreme Court wrote not once—but twice in *TransUnion*. "Central to assessing concreteness is whether the asserted harm has a 'close relationship' to a harm 'traditionally' recognized as providing a basis for a lawsuit in American courts—such as physical harm, monetary harm, or various intangible harms including…reputational harm." *Id.* at 417. The allegations added to Plaintiffs' Proposed First Amended Complaint do not confer standing on Plaintiffs to pursue either damages or injunctive relief.

As to Plaintiffs' claim for damages, Plaintiffs assert the *same* injury in their Proposed First Amended Complaint as they did in their initial complaint. The "asserted injury" is still the unauthorized removal of CMI from their copyright-protected work. Plaintiffs still do not plausibly allege that a copy of their work from which the CMI has actually been removed has been disseminated by ChatGPT to anyone in response to any specific inquiry.

---

[1] I am aware that this case is among a group of cases that are being transferred to my colleague, The Hon. Sidney Stein, by the Judicial Panel on Multi-District Litigation (MDL). To the best of my knowledge the MDL Panel's order has not yet been filed, so the case remains on my docket. I have spoken with Judge Stein, and advised him that we were about to issue this opinion. He asked that I do so if the decision would be completed prior to the moment when the case was transferred to him. Given that the pending motion is a motion for leave to amend following my dismissal of the complaint, it is particularly appropriate that I finish my work.

Plaintiffs principally add details on the technical process by which OpenAI uploaded Plaintiffs' copyright-protected works into their training sets—a process that was internal to OpenAI. These additional allegations do not render Plaintiffs' asserted harm more concrete. If Plaintiffs believe that I got it wrong in my previous order, and that my esteemed colleague Judge Rakoff got it right—that Plaintiffs' asserted injury does have a "close historical or common-law analogue" and that analogue is copyright infringement, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-CV-1515 (JSR), 2025 WL 556019 (S.D.N.Y. Feb. 20, 2025)—then this motion is not the place to make that argument. Seek review before Second Circuit.

Plaintiffs' alternate analogue of unjust enrichment is further from the mark. Their argument, as far as I understand it, is that by stripping Plaintiffs' CMI before uploading articles into ChatGPT, OpenAI saved itself the time, manhours, expertise and other costs that would have been required to *better* train ChatGPT to omit CMI when outputting an answer that integrated—but did not plagiarize—Plaintiffs' work. Plaintiffs do not explain how either the benefit or the expense of those costs would have found its way to or from Plaintiffs' pockets. As such, I don't find this to be the historical analogue for Plaintiffs' asserted injury.

Turning to the next issue, Plaintiffs still lack standing for injunctive relief. This was going to be a tall order. "Given the quantity of information contained in the repository [of ChatGPT], the likelihood that ChatGPT would output plagiarized content from one of Plaintiffs' articles seems remote." Dkt. No. 117, at 9. Instead of adding new evidence of any content output by ChatGPT that was plagiarized from Plaintiffs' works, Plaintiffs have amended their complaint to include the allegations of plagiarism made by *other* litigants in *other* ongoing lawsuits against OpenAI. Dkt. No. 118, ¶¶ 65, 67-68. While these allegations may move the needle, they are insufficient to meet

the high threshold of "certainly impending" injury or "substantial risk" that the harm will occur. *See Susan B. Anthony List v. Driehaus,* 573 U.S. 149, 157 (2014).

For the very same reason, I will not grant Plaintiffs' request to proceed with jurisdictional discovery. Plaintiffs are looking for a needle in a haystack. "A plaintiff is not… entitled to jurisdictional discovery if it cannot show that the requested discovery is likely to produce the facts needed to withstand a Rule 12(b)(1) motion. *Molchatsky v. U.S.*, 778 F. Supp. 2d 421, 438 (S.D.N.Y. 2011) (internal quotations and citations omitted).

CONCLUSION

For the reasons stated above, Plaintiffs' Motion for Leave to Amend the Complaint, or in the Alternative, to Continue Taking Jurisdictional Discovery is DENIED.

This constitutes a written opinion. The Clerk is directed to remove the motions at Dkt. No. 118 from the Court's list of open motions.

Dated: April 3, 2025

U.S.D.J.

BY ECF TO ALL COUNSEL

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-SHS

[PROPOSED] ORDER

The *Raw Story* Plaintiffs' Motion for Reconsideration is GRANTED. The *Raw Story* Plaintiffs shall file their Proposed First Amended Complaint within seven days of this Order.

IT IS SO ORDERED

Dated: _____

_____
Honorable Sidney H. Stein
United States District Court Judge