UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

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

PLAINTIFF'S COMBINED MEMORANDUM OF LAW IN OPPOSITION TO MICROSOFT'S AND OPENAI DEFENDANTS' MOTIONS TO DISMISS

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page

Cases

Statutes

Other Authorities

I. INTRODUCTION

Defendants would brush this case aside for raising no "critical legal questions … about the application of longstanding copyright principles to new technology." OpenAI Mot. at 1. That has no bearing on whether The Intercept has stated a claim, but it is also wrong: the Digital Millennium Copyright Act was passed 25 years ago to better protect copyright owners from infringement through the emerging internet, and the Court is called up on in this case to apply the DMCA to another technological evolution that threatens human content creators.

After conceding that their products exist only by being "fe[d] large amounts of text" written by people and put through some undisclosed "algorithm," Microsoft Mot. at 1, Defendants claim that The Intercept fails to state DMCA claims and lacks standing to bring them. On standing, Defendants argue that The Intercept must allege specific works that ChatGPT disseminated to the public. But DMCA standing does not require dissemination. As with copyright infringement (the closest historical analogue to the DMCA), the injury is the interference with a plaintiff's right to exclude others from using its copyrighted works irrespective of dissemination. The Intercept has alleged just that. Plus, The Intercept has alleged dissemination between OpenAI and Microsoft.

On Rule 12(b)(6), Defendants have hidden much of the content of their training sets and now attempt to use their own secrecy as a basis for dismissal, while demanding far more of The Intercept's Complaint than the pleading stage requires and ignoring many of the Complaint's allegations. Defendants are sufficiently on notice that The Intercept's claims are based on the works published on The Intercept's website and placed into Defendants' training sets with author, title, copyright notice, and/or terms of use information removed, and Defendants cannot and do not claim ignorance of the contents of their own training sets. And while Defendants argue, disingenuously and incorrectly, that The Intercept was required to identify specific instances of

regurgitation of The Intercept's works, when the New York Times made such allegations, Defendants accused them of nothing less than computer hacking.[1]

The Intercept has adequately pled scienter as well. The Second Circuit allows for lenient scienter pleading, and even if a greater level of detail were required, The Intercept has pled abundant facts, well beyond the bare-bones pleadings that led to dismissals in some out-of-jurisdiction DMCA cases—including that ChatGPT plagiarizes substantial content. Defendants' contrary arguments rest on supposed pleading rules this District has expressly rejected. *Compare, e.g.*, OpenAI Mot. at 16 (arguing for dismissal because The Intercept did not allege "when, why, or how" the violations occurred) *with Pilla v. Gilat*, No. 19-cv-2255, 2020 WL 1309086, at *12 (S.D.N.Y. Mar. 19, 2020) ("Although Plaintiff does not allege how, when, or where this removal occurred, such details are not necessary at the pleading stage for a claim under the DMCA.").

Much like the hallucinations to which their products are prone,[2] Defendants ignore or mischaracterize both the applicable law and the allegations of the Complaint. The Intercept is entitled to its day in court. At most for Defendants, the Court should permit The Intercept to amend to address any deficiencies.

II. BACKGROUND

A. The Intercept publishes thousands of news articles online that contain CMI.

The Intercept is an award-winning news organization. Compl. ¶ 8. Its articles are published on the internet. *Id*. ¶ 32; *see also* www.theintercept.com. At the time of publication, its articles are conveyed with author, title, copyright, and terms of use information. Compl. ¶ 32.

---

[1] Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss, 2, *The New York Times Company v. Microsoft Corp.*, No. 23-cv-11195 (S.D.N.Y. Feb. 26, 2024).

[2] *See, e.g.*, Molly Bohannon, Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge Is Considering Sanctions (Forbes June 8, 2023), https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/?sh=34a922fa7c7f.

B. Defendants intentionally remove CMI from The Intercept's news articles.

Generative AI systems and large language models, including ChatGPT, are trained on works created by humans. Compl. ¶¶ 4-5. Once trained, an LLM is able to provide responses to user prompts. *Id.* ¶ 34. These responses sometimes mimic material from the works on which they are trained or regurgitate those works entirely. *Id.* ¶¶ 35-36. When that happens, ChatGPT generally does not provide the author, title, copyright notice, or terms of use information contained in the original version of the work. *Id.* ¶ 40.

Beginning with GPT-4, Defendants have hidden the precise contents of the training sets on which their products are built. *Id.* ¶ 29. But information exists about prior ChatGPT training sets, and that information shows that Defendants have trained their products on thousands of The Intercept's copyright-protected news articles. *Id.* ¶ 38. In particular, Defendant trained ChatGPT on training sets called WebText and WebText2—sets created by OpenAI that are collections of links posted on the website Reddit. *Id.* ¶¶ 30-31. Defendants also created training sets derived from a repository called Common Crawl, which is a "scrape of most of the internet" created by a third party. *Id.* Defendants have not published the contents of WebText, WebText2, or their training sets derived from Common Crawl. *Id.* ¶ 29. But various public sources have recreated approximations of these sets. *Id.* ¶ 38. And in those approximations, thousands of The Intercept's articles appear without the copyright management information with which The Intercept conveyed them to the public. *Id.* There is only one plausible explanation: Defendants intentionally removed CMI as part of their LLM training. After all, given the nature of LLM training, if ChatGPT had been trained on works that included CMI, it would have learned to output CMI. *Id.* ¶ 39.

Likewise, Defendants knew, or had reasonable grounds to know, that their removal of The Intercept's CMI in their training sets would likely further or conceal infringement by both themselves and ChatGPT users. Indeed, OpenAI recently created tools to allow copyright owners

to block their work from being incorporated into training sets and has reached licensing deals with some media organizations, suggesting (at least in the light most favorable to The Intercept) that it knows that copying journalists' works is likely infringement. *Id*. ¶¶ 62, 63. The removal of CMI, in turn, furthers and conceals Defendants' infringement at least by preventing ChatGPT users from knowing that outputs are based on copyright-protected works of journalism. *Id*. ¶ 50. It also furthers ChatGPT users' infringement at least by encouraging them to distribute outputs the users do not know are infringing. *Id*. ¶¶ 48-49. And it facilitates Defendants' large-scale copying and use of copyright-protected material in their training sets by avoiding the problems for their products that would arise if Defendants had included CMI. *Id.* ¶¶ 34, 39, 50.

C. Defendants intentionally distribute CMI-less works to each other.

OpenAI and Microsoft have a close working relationship: Microsoft has invested billions of dollars in OpenAI and will have a 49% stake in the company after its investment has been repaid. Compl. ¶¶ 20. As part of that relationship, Microsoft provides the data center and supercomputing infrastructure used to train ChatGPT, and hosts ChatGPT training sets. *Id*. ¶ 21. In connection with the development of ChatGPT, OpenAI and Microsoft have shared CMI-less copies of The Intercept's copyright-protected news articles with each other. *Id*. ¶¶ 45-46, 65, 77. And given their removal of the CMI, Defendants knew that those articles lacked CMI.

III. LEGAL STANDARDS

On a motion to dismiss under Rule 12(b)(1) for lack of standing, the court's resolution depends on whether the motion is facial—based solely on the allegations in the complaint—or fact-based. *Sonterra Cap. Master Fund Ltd. v. UBS AG*, 954 F.3d 529, 533 (2d Cir. 2020). When the motion is facial, as Defendants' motions are here, the court must "accept[] as true all material factual allegations of the complaint, and draw[] all reasonable inferences in favor of the plaintiff." *Id*. (cleaned up). In these cases, "the plaintiff has no evidentiary burden." *Id*.

Likewise, under Rule 12(b)(6), the court must "accept[] all factual allegations as true and draw[] all reasonable inferences in favor of the plaintiff." *Sierra Club v. Con-Strux, LLC*, 911 F.3d 85, 88 (2d Cir. 2018). The court must deny the motion if the complaint "contain[s] sufficient factual matter, accepted as true, to 'state a claim that is plausible on its face.'" *Id*. (quoting *Ashcroft v. Iqbal*, 556 U.S. 662, 678, (2009)). And because the purpose of a complaint is to "give the defendant fair notice of what the ... claim is and the grounds upon which it rests," a plaintiff need not plead "specific facts" in order to overcome a motion to dismiss. *Erickson v. Pardus*, 551 U.S. 89, 93 (2007) (cleaned up) (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 545 (2007)). Defendants may dispute the accuracy of The Intercept's allegations or question The Intercept's ability to prove them, but those are issues for trial, not the pleading stage. *See Giuffre v. Dershowitz,* 410 F. Supp. 3d 564, 577 (S.D.N.Y. 2019) ("[P]laintiffs can put their allegations out to the world and must only plead them, not prove them, at the motion to dismiss stage.").

IV. ARGUMENT

A. The Intercept has Article III standing.

Article III standing requires, *inter alia*, a "concrete and particularized" injury. *Spokeo, Inc. v. Robins*, 578 U.S. 330, 339 (2016). The Intercept plausibly alleges both.

1. The Intercept's injuries are concrete.

An injury is concrete if it has a "close historical or common-law analogue." *TransUnion LLC v. Ramirez*, 594 U.S. 413, 424 (2021). The analogue need not be an "exact duplicate." *Id*. at 433. Instead, "some relationship to a well-established common-law analog" will do. *Bohnak v. Marsh & McLennan Cos., Inc*., 79 F.4th 276, 285 (2d Cir. 2023); *see also Saba Cap. Cef Opportunities 1, Ltd. v. Nuveen Floating Rate Income Fund*, 88 F.4th 103, 115-16 (2d Cir. 2023) (holding that dilution of voting shares is analogous to a common-law "property-based injury"). In deciding concreteness, "[c]ourts must afford due respect to Congress's decision to impose a

statutory prohibition or obligation on a defendant, and to grant a plaintiff a cause of action to sue over the defendant's violation of that statutory prohibition or obligation." *TransUnion*, 594 U.S. at 425. Both "tangible" and "intangible" harms are concrete. *See id*. at 424-425.

The unlawful removal of CMI from a copyright-protected work (Counts I and III), and the distribution of CMI-less copies of those works (Counts II and IV), are analogous to copyright infringement. Congress evidently saw the two as analogous: it called the DMCA an Act "[t]o amend title 17, United States Code," which exclusively concerns copyright. Pub. L. 105-304, 112 Stat. 2860 (1998). It did so because CMI protects the integrity of copyrighted works. *See* S. Rep. 105-190 at 16 (1998). Further recognizing the analogy, Congress provided similar remedies for DMCA violations as it long had for copyright infringement: in both, the plaintiff can choose between actual damages and profits on the one hand, and statutory damages on the other. *Compare* 17 U.S.C. §§ 504(b), (c) (copyright infringement) *with* 17 U.S.C. § 1203(c) (DMCA violations)). While not dispositive, Congress's view on the matter is entitled to considerable weight. *See Spokeo*, 578 U.S. at 341 ("[B]ecause Congress is well positioned to identify intangible harms that meet minimum Article III requirements, its judgment is also instructive and important.").

The analogy between copyright infringement and DMCA violations also follows from first principles: both the Copyright Act and the DMCA protect similar rights involving copyright-protected works. The Copyright Act protects certain exclusive rights, such as the rights to reproduce the work, prepare derivative works, and distribute the work. *See* 17 U.S.C. § 106 (listing exclusive rights). The DMCA grants copyright owners similar rights. In particular, the protection against removing or altering CMI, 17 U.S.C. § 1202(b)(1), is analogous to the rights to reproduce the works and prepare derivative ones, 17 U.S.C. §§ 106(1), (2): both grant the copyright owner the sole prerogative to decide how future iterations of the work may differ from the version the

owner published. And the protection against distributing CMI-less works, 17 U.S.C. § 1202(b)(3), is analogous to the exclusive right to distribute copies of the work, 17 U.S.C. § 106(3): both allow the copyright owner to decide who sees the work and in what form.

Given this analogy, The Intercept has alleged a concrete injury: Defendants' interference with its exclusive right to control its copyrighted works by removing CMI from them and distributing them in a CMI-less form without authority. *See* Compl. ¶¶ 51-77. For copyright infringement, courts have never required more. *See Fox Film Corp. v. Doyal*, 286 U.S. 123, 127 (1932) (describing copyright owner's right as one simply to "exclude others from using his property"). And this accords with the common law, which recognizes interference with property, without more, as a concrete injury. *See* Restatement (Second) of Torts § 163 ("One who intentionally enters land in the possession of another is subject to liability to the possessor for a trespass, although his presence on the land causes no harm to the land, its possessor, or to any thing or person in whose security the possessor has a legally protected interest."). Given the infringement-DMCA analogy, the same holds for the latter: the unlawful removal of CMI from a copyrighted work, or distribution of such a work without CMI, is a concrete injury.

Defendants would impose two more conditions for standing. First, they would require economic harm. *See* OpenAI Mot. at 5. But standing does not require this. *See TransUnion*, 594 U.S. at 425. And neither, historically, has copyright infringement—the relevant analogy. This is clear from the 1790 version of the Copyright Act, passed by the first Congress, which granted statutory damages of 50 cents per infringing page without any further showing. *See* Act of May 31, 1790, ch. 15, § 2. The same rule has persisted, with the Supreme Court making clear long ago that liability may lie "[e]ven for uninjurious and unprofitable invasions of copyright." *F. W. Woolworth Co. v. Contemp. Arts*, 344 U.S. 228, 233 (1952); *see also Jewell-La Salle Realty Co.*

v. Buck, 283 U.S. 202, 208 (1931) (construing Copyright Act to mandate minimum statutory damages of $10 per performance even if "there is no showing as to actual loss"); 17 U.S.C. § 504(c)(1) (providing for statutory damages for copyright infringement without regard to economic loss). Given the analogy between DMCA violations and copyright infringement, the same result follows: DMCA violations do not require economic harm. Defendants cite no contrary authority.

Second, Defendants would require that defendants disseminate the plaintiff's works. *See* Microsoft Mot. at 9-10; OpenAI Mot. at 5-7. But copyright infringement—the relevant historical analogue—has never required dissemination. This has been true from the 1790 Copyright Act to the present version of the law. *See* Act of May 31, 1790, ch. 15, § 2 (imposing liability on anyone who "shall print, reprint, publish, or import" a copyrighted work); *A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1014 (9th Cir. 2001) (holding that downloading of files containing copyrighted music violates the reproduction right). Given the close analogy between copyright infringement and DMCA violations, nothing justifies treating DMCA violations differently. And even if there were, Defendants' arguments would fail on their own terms: The Intercept alleges that Microsoft and OpenAI distributed the works to each other. *See* Compl. ¶¶ 64-65; 76-77.

Microsoft contends that *TransUnion* supports a dissemination requirement. Microsoft Mot. at 8-10. But it is far afield. Its plaintiffs alleged that a credit reporting agency did not keep accurate credit files. *See TransUnion*, 594 U.S. at 421. Likening their injury to one at common law, the plaintiffs chose defamation. *See id*. at 432. The Court held that the analogy justified a finding of injury for plaintiffs whose credit files were disseminated to third parties, but not for those whose files were not. *See id*. at 433. It reached that conclusion because defamation requires publication. *See id*. at 434. Thus, its dissemination requirement was an artifact of the plaintiffs'

chosen analogy to a historical injury that requires it. It has no bearing on a case like this, where The Intercept analogizes to a different historical injury—copyright infringement—that does not.[3]

2. The Intercept's injuries are particularized.

"For an injury to be 'particularized,' it 'must affect the plaintiff in a personal and individual way.'" *Spokeo*, 578 U.S. at 339 (quoting *Lujan v. Defs. of Wildlife*, 504 U.S. 555, 560 n.1 (1992)). The Intercept met that requirement by alleging that Defendants removed CMI from ***its*** copyright-protected news articles and distributed ***its*** articles knowing that they lacked CMI. *See* Compl. ¶¶ 51-77. If such removal and distribution constitute Article III injuries—and they do for the reasons just given—then the Intercept has alleged that it suffered them in a personal and individual way.

3. *Doe 1* supports standing.

Defendants' standing arguments rely heavily on *Doe 1 v. Github*, 672 F. Supp. 3d 837 (N.D. Cal. 2023). *Doe 1* rejected standing for damages but found standing for an injunction. The case is readily distinguished, and otherwise supports standing here.

On damages, *Doe 1* held that plaintiffs did not allege a particularized injury because they did not allege dissemination of their own works. *See id*. at 850. But like *TransUnion*, it required dissemination only because of how plaintiffs defined injury: as a violation of their licenses, which forbade CMI-less dissemination. *See id*. Because plaintiffs did not allege CMI-less disseminations of their own works, they "do not allege that they themselves have suffered the injury they describe," and thus failed particularity. *Id*. The Court offered no view on the analogy to copyright infringement because it was not asked to. Its holding on standing for damages is inapt.

[3] Microsoft also cites cases regarding the right of attribution and argues that they are not analogous. *See* Microsoft Mot. at 10. Those are irrelevant. The Intercept is not analogizing its injury to the injuries in those cases, and the plaintiff gets to pick its analogy. *See TransUnion*, 594 U.S. at 424 (holding that the inquiry "asks whether plaintiffs have identified a close historical or common-law analogue for their asserted injury").

But *Doe 1*'s holding on standing for an injunction is (potentially) on point. It held, on a dissemination-based theory, that the plaintiffs had standing by alleging "a substantial risk that Defendants' programs will reproduce Plaintiffs' licensed code as output." *Id*. at 851. The plaintiffs had alleged that the programs were trained on their source code, that the programs sometimes reproduced well-known code (though not plaintiffs' own), and that one of the programs reproduces code "about 1% of the time." *Id*. The Intercept has easily cleared that bar. It alleges that ChatGPT was trained on The Intercept's copyrighted works, Compl. ¶ 38, that ChatGPT has reproduced copyrighted works of journalism, *id*. ¶¶ 35-36, and that "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." Compl. ¶ 5. This conveys a much greater risk than 1%.

Plus, *Doe 1* undermines any argument that plaintiffs must plead dissemination of their ***own*** works to have standing to pursue an injunction: *Doe 1* allowed plaintiffs' injunction claim to proceed after concluding that plaintiffs had ***not*** alleged any regurgitations of their own works. *See Doe 1*, 672 F. Supp. 3d at 850-51.[4] Thus, even if standing did require dissemination, *but see* Section IV.A.1, The Intercept has plausibly alleged facts to support standing for an injunction.

Finally, Microsoft argues that, in amending their complaint, the *Doe 1* plaintiffs "pleaded themselves out of their Section 1202(b)(1) and 1202(b)(3) claims" by alleging that Microsoft Copilot did not produce identical copies of their works. *Doe 1 v. GitHub, Inc*., No. 22-cv-06823,

[4] Citing *Clapper v. Amnesty International USA*, 568 U.S. 398, 414 (2013), Microsoft argues that The Intercept has no standing for a dissemination-based injury because dissemination depends on independent actors. *See* Microsoft Mot. at 12. This argument has no force against The Intercept's non-dissemination-based injury. But as *Doe 1* held, it also fails if injunctive relief required the risk of dissemination: unlike *Clapper*, the risk of dissemination is substantial as a matter of fact. *See Doe 1 v. GitHub, Inc*., No. 22-cv-06823, 2024 WL 235217, at *4 (N.D. Cal. Jan. 22, 2024).

2024 WL 235217, at *8 (N.D. Cal. Jan. 22, 2024). It then extrapolates—from the pleading decisions of a single set of plaintiffs—a supposed "proposition that bare allegations that unidentified works were included ***in a training set*** for an AI tool are insufficient to allege that the tool has or will cause injury." Microsoft Mot. at 13 (emphasis in original). This argument is based on a flawed reading of *Doe 1*. *Doe 1*'s analysis has nothing to do with standing; at that point the court had already ***found*** standing. *See Doe 1*, 2024 WL 235217, at *5. Instead, the court held that the amended complaint failed to state a claim under Rule 12(b)(6) by not alleging identicality. *See id*. at *6, *8. Regardless, *Doe 1* issued no grand pronouncement that LLMs do not produce identical copies of works. Its holding rests entirely on unproven allegations by a different plaintiff in a different case—allegations not binding here: "Other cases presenting different allegations and different records may lead to different conclusions." *Twitter, Inc. v. Taamneh*, 598 U.S. 471, 507 (2023) (Jackson, J., concurring). The Intercept has Article III standing.

B. The Intercept has statutory standing.

OpenAI argues that The Intercept lacks "statutory standing" because it supposedly "is not within the class of plaintiffs that Congress authorized to sue" for DMCA violations—a class OpenAI does not define. OpenAI Mot. at 7, 9. *See* 17 U.S.C. § 1203(a) ("Any person injured by a violation of section 1201 or 1202 may bring a civil action."). This argument fails for the same reason as the last: The Intercept has suffered an Article III injury, and OpenAI gives no reason to believe that "injury" under section 1203(a) means anything different than it does under Article III.

In a similar vein, Microsoft argues that "[r]emoval of CMI without dissemination of a copy of a work … does not result in even the sort of injury Congress had in mind when it enacted the DMCA," which harms it casts as those flowing from "public dissemination" of that work without CMI. Microsoft Mot. at 11. But it cites no case inferring its conclusion from the supposed purpose of the DMCA. Indeed, its argument improperly rests entirely on snippets from a Senate report and

does not engage with the text or structure of section 1202(b). *See Associated Press v. All Headline News Corp*., 608 F. Supp. 2d 454, 461 (S.D.N.Y. 2009) (observing in a DMCA case that "the Second Circuit has held that legislative history should not be considered as a first resort, and that statutory language should be applied as written");[5] *see also Food Mktg. Inst. V. Argus Leader Media*, 588 U.S. 427, 437 (2019) (overruling long line of precedent that "inappropriately resort[ed] to legislative history before consulting the statute's text and structure").

Microsoft's interpretation contradicts text and structure. Section 1202(b) creates three violations: removing or altering CMI, distributing false or removed CMI, and distributing works knowing that CMI has been removed or altered. *See* 17 U.S.C. §§ 1202(b)(1), (2), (3). Congress chose to make dissemination an element of the second and third violations but not the first. Thus, it evidently determined that removal or alteration itself constitutes a harm under section 1202(b) whether or not the violator also disseminated the work. Requiring dissemination for a section 1202(b)(1) removal claim would render that provision essentially duplicative of section 1202(b)(3): anyone who intentionally removed CMI from a work in violation section 1202(b)(1), and then disseminated the work, would necessarily have distributed it knowing that CMI had been removed, thus violating section 1202(b)(3). *See Argus Leader*, 588 U.S. at 427.

Section 1202(b)(1) does not require dissemination, and The Intercept alleged it for section its 1202(b)(3) claims.[6] The Intercept has alleged an injury under section 1203(a).

---

[5] The legislative history Microsoft cites also does not say that injury requires dissemination. *See* S. Rep. 105-190 at 11 n.18 (1998); *id*. at 16.

[6] OpenAI's cases on section 1203's injury requirement says nothing about dissemination. They that plaintiff's asserted injury of "confusion in the marketplace" did not qualify because defendant did not participate in that market, *Alan Ross Mach. Corp. v. Machinio Corp*., No. 17-cv-3569, 2019 WL 1317664, at *4 (N.D. Ill. Mar. 22, 2019), and that a DMCA violation did not cause the plaintiff to lose a separate copyright case, *Steele v. Bongiovi*, 784 F. Supp. 2d 94, 97-98 (D. Mass. 2011). Microsoft's cases say nothing about injury or section 1203. *See generally Fashion Nova,*

C. The Intercept states a claim under section 1202(b)(1).

A violation of section 1202(b)(1) requires "(1) the existence of CMI on the allegedly infringed work, (2) the removal or alteration of that information and (3) that the removal was intentional." *Fischer v. Forrest*, 968 F.3d 216, 223 (2d Cir. 2020). It also requires defendant to know, or have "reasonable grounds to know," that removal "will induce, enable, facilitate, or conceal" infringement. 17 U.S.C. § 1202(b)(1). The Intercept alleges each element.

1. The Intercept alleges the existence and removal of CMI from its articles.

As to the first and second elements, The Intercept has alleged both the existence of CMI on its works and that Defendants removed it. *See* Compl. ¶¶ 32, 38-44. Defendants collectively resist this conclusion on four grounds, though OpenAI only endorses one.

First, Microsoft says that The Intercept did not identify what type of CMI was removed from its news articles. *See* Microsoft Mot. at 17-18. It analogizes The Intercept's allegations to those in *Andersen v. Stability AI Ltd*., No. 23-cv-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023), which dismissed a complaint for not "identify[ing] the particular types of their CMI from their works that they believe were removed or altered." *Id*. at *11. But that ignores the Complaint, which alleges that Defendants removed author, title, copyright notice, and terms of use from its works. *See, e.g.*, Compl. ¶¶ 38-44. These are all types of CMI. *See* 17 U.S.C. § 1202(c) (defining "copyright management information" to include all four). This charge is insubstantial.

Second, Defendants argue that The Intercept did not name the specific works from which they removed its CMI. They argue that such allegations are required to state a claim. *See* Microsoft Mot. at 15, 17-18; OpenAI Mot. at 10-12. But The Intercept has identified its works

---

LLC v. Blush Mark, Inc., No. 22-cv-6127, 2023 WL 4307646 (C.D. Cal. June 30, 2023); *Roberts v. BroadwayHD* LLC, 518 F. Supp. 3d 719 (S.D.N.Y. 2021).

with the required level of detail in the context of this case. As OpenAI admits, the purpose of identifying a work is to "give the defendants fair notice of the claims against them." OpenAI Mot. at 11 (quoting *Dow Jones & Co., Inc. v. Int'l Sec. Exch., Inc.*, 451 F.3d 295, 307 (2d Cir. 2006)). The Intercept has alleged removal of CMI from its works contained in Defendants' training sets. *See* Compl. ¶ 42. And Defendants cannot seriously claim ignorance of what works those are. OpenAI has "published a list of the top 1,000 domains present in WebText and their frequency,"[7] which notes that WebText—one of the training sets referenced in the Complaint—contains exactly 6,484 (unidentified) URLs from The Intercept's web domain.[8] Thus, it can isolate The Intercept's news articles in its training sets. That suffices to notify Defendants of the works at issue.[9]

Further, The Intercept cannot name all its works contained in Defendants' training sets only because Defendants have kept them secret. *See* Compl. ¶ 29. Requiring a plaintiff to name all its works in these circumstances—where, due to defendant's actions, only it knows what the works are—would perversely incentivize defendants to conceal their DMCA violations. And it would do so without advancing the principle animating the identification requirement: to provide defendants notice of the claims against them. If Defendants know the works, they know the claims.

In none of Defendants' laundry list of cases, unlike here, could the defendant actually identify the works at issue given the plaintiff's allegations and information in its sole possession. For instance, in this District's first case to require identification, the complaint alleged that defendant infringed a work called "The Skyrider," plaintiff owned two works containing the word

[7] GPT-2 model card (Nov. 2019), https://github.com/openai/gpt-2/blob/master/model_card.md.
[8] Domains.txt (2019), https://github.com/openai/gpt-2/blob/master/domains.txt.
[9] Though not specifically alleged, The Intercept's claims only relate to articles published on its web domain, theintercept.com, which Defendants are able to identify. Further, to the extent that the copyrights in a limited number of articles are owned by their freelance authors, these can easily be addressed in discovery. If the Court prefers, The Intercept can replead to address this.

"Skyrider," and it was unclear to which work the complaint referred. *Calloway v. Marvel Ent. Grp., a Div. of Cadence Indus. Corp*., No. 82-cv-8697, 1983 WL 1141, at *3-*4 (S.D.N.Y. June 30, 1983). The Intercept's complaint contains no such ambiguity. And in the other cases that found the works insufficiently described—only one of which involved the DMCA—plaintiffs not only did not specifically name the works but also did not describe the works in a way that enabled defendants to identify them. *See Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1175 (N.D. Cal. 2019) (plaintiff alleged CMI removal "without providing any facts to identify which photographs had CMI removed"); *see also Palmer Kane LLC v. Scholastic Corp*., No. 12-cv-3890, 2013 WL 709276, at *3 (S.D.N.Y. Feb. 27, 2013) (plaintiff provided list of some works and indicated without elaboration that "this list is not exhaustive"); *Wolo Mfg. Corp. v. ABC Corp*., 349 F. Supp. 3d 176, 201 (E.D.N.Y. 2018) (plaintiff described the allegedly infringed works simply as "Other Works"); *Cole v. John Wiley & Sons, Inc*., No. 11-cv-2090, 2012 WL 3133520, at *12 (S.D.N.Y. Aug. 1, 2012) (for one defendant, plaintiff listed two works and alleged "that the claim is also intended to cover other, unidentified works"; for the others, plaintiff listed works but did not allege that defendants infringed any); *Joint Stock Co. Channel One Russia Worldwide v. Infomir LLC*, No. 16-cv-1318, 2017 WL 696126, at *14 (S.D.N.Y. Feb. 15, 2017) (plaintiff did not specify representative sample of the works, which in that case's context prevented defendants from evaluating the other elements); *Sherwood 48 Assocs. v. Sony Corp. of Am*., 76 F. App'x 389, 391 (2d Cir. 2003) (plaintiff failed to precisely describe trade dress).[10] Thus, in the context of this

[10] Defendants' other cases did not dismiss the relevant complaint for failing to identify the work. *See Dow Jones*, 451 F.3d at 307 (plaintiff identified the marks but did not provide "any factual allegations concerning the nature of the threatened use"); *Felix the Cat Prods., Inc. v. Cal. Clock Co*., No. 04-cv-5714, 2007 WL 1032267, at *4 (S.D.N.Y. Mar. 30, 2007) (plaintiff had identified the work but court dismissed complaint because it failed to specify other details concerning the infringement); *Sitnet LLC v. Meta Platforms, Inc*., No. 23-cv-6389, 2023 WL 6938283, at *1 (S.D.N.Y. Oct. 20, 2023) (adopting additional provisions for case management plan).

case, The Intercept has done all it needs to put Defendants on notice of its claims, especially since the claims are the same for all of the works: Defendants took copies of works published by The Intercept on the internet without permission, stripped away copyright management information, and used them to train their products with the requisite intent and knowledge.

Third, Microsoft (but, again, not OpenAI) argues that The Intercept has not pled enough facts to plausibly allege that Defendants removed CMI from its works. It says that The Intercept has not identified the publicly available information that led it to plead that "thousands of Plaintiff's copyrighted works were included in Defendants' training sets." Microsoft Mot. at 15 (quoting Compl. ¶ 42). That is just wrong. The Complaint identifies that information as public approximations of the ChatGPT training sets—the best any plaintiff could do given that the actual training sets are all secret—that contain "[t]housands of Plaintiff's works" without author, title, copyright notice, and terms of use information. Compl. ¶ 38. Microsoft cites no case suggesting The Intercept was required to allege more, and indeed it was not. *See Erickson*, 551 U.S. at 93 (holding that plaintiff need not allege "specific facts" to overcome a motion to dismiss).

Microsoft next claims that The Intercept has not alleged that CMI was removed from "***its own*** works" as opposed to other works in the ChatGPT training sets. Microsoft Mot. at 16 (emphasis in original). But this, too, ignores the allegation that the recreated approximations of the ChatGPT training sets contain thousands of "Plaintiff's works" lacking CMI. Compl. ¶ 38.

Microsoft further argues that the removal allegation depends on a "daisy-chain of hypotheticals" about how ChatGPT's outputs depend on its training data. Microsoft Mot. at 15-16. Two responses. First, for reasons just discussed, the Complaint plausibly alleges removal without regard to the relation between inputs and outputs: it is plausible that Defendants removed CMI given that The Intercept's works contained CMI when published and the same types of CMI

are absent from public approximations of the ChatGPT training sets. *See* Compl. ¶¶ 32, 38. Second, it is plausible that an LLM doesn't output CMI precisely because it wasn't trained on CMI. Other explanations (which Microsoft does not offer) might be available, but The Intercept need not disprove them at the pleading stage. *See George & Co. LLC v. Target Corp.*, No. 21-cv-4254, 2022 WL 1407236, at *12 (E.D.N.Y. Jan. 27, 2022) (holding that a complaint need not "disprove an obvious alternative explanation" to satisfy *Twombly*) (cleaned up).

Fourth, Microsoft argues that The Intercept has not plausibly alleged that any CMI removal was performed by Microsoft, as opposed (or in addition) to OpenAI. Microsoft Mot. at 18-20. The crux of The Intercept's allegations here is the close working relationship between Microsoft and OpenAI in the development of ChatGPT—that Microsoft "has invested billions of dollars in OpenAI Global LLC and will own a 49% stake in the company after its investment has been repaid," and that Microsoft "provides the data center and supercomputing infrastructure used to train ChatGPT." Compl. ¶¶ 20-21. Given that relationship, it is at least plausible that Microsoft participated with OpenAI in the removal of The Intercept's CMI. And contrary to Microsoft's supposition, The Intercept's allegation is not that Microsoft is a mere passive host of OpenAI's training data. It is that, given the close working relationship between the two, Microsoft is plausibly an active participant in the development of the ChatGPT training sets and the corresponding removal of The Intercept's CMI.[11]

Finally, Microsoft says that the Complaint "raises more questions than it answers" regarding CMI-less outputs—in particular, how outputs could be expected to provide the author's

[11] Microsoft expresses confusion about The Intercept's allegations that it and OpenAI created "Common Crawl datasets" given that Common Crawl itself originated elsewhere. *See* Compl. ¶ 31. But the Complaint's allegations are clear: Defendants took data from Common Crawl, which is a "scrape of most of the internet," *id.* ¶ 30, and created training sets from that data.

byline if they only emit "significant amounts" of material from copyright-protected works. Microsoft Mot. at 16 (quoting Compl. ¶ 36). It raises such "questions," Microsoft says, because CMI-less excerpts supposedly cannot give rise to a DMCA claim. But that is incorrect. *See Planck LLC v. Particle Media, Inc.*, No. 20-cv-10959, 2021 WL 5113045, at *6 (S.D.N.Y. Nov. 3, 2021) (holding that plaintiff "states a claim for a violation of § 1202(b)" by alleging that "Defendants have distributed infringing excerpts from Plaintiff's stories" without CMI); *Doe 1*, 672 F. Supp. 3d at 857 (holding plaintiffs stated DMCA claims by alleging that defendants' technology "excerpts code without the accompanying CMI"). Plus, the Complaint also alleges that ChatGPT has provided responses that "regurgitate verbatim or nearly verbatim copyright-protected works of journalism" without CMI, where one would necessarily expect a byline. Compl. ¶ 35.

2. The Intercept alleges scienter.

A section 1202(b)(1) plaintiff must allege that the defendant removed the CMI both intentionally and while knowing, or with 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)(1). Courts call this requirement "double-scienter." *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 171 (2d Cir. 2020). The "scienter" label is significant because "[t]he Second Circuit has stated that courts should be lenient in allowing scienter issues to survive motions to dismiss." *Aaberg v. Francesca's Collections, Inc.*, No. 17-cv-115, 2018 WL 1583037, at *9 (S.D.N.Y. Mar. 27, 2018) (citing *In re DDAVP Direct Purchaser Antitrust Litig.*, 585 F.3d 677, 693 (2d Cir. 2009)). Thus, courts in this District have allowed DMCA claims to proceed based even on "sparse" allegations of scienter, including where plaintiff alleged only that defendant "'intentionally and knowingly removed copyright information identifying Plaintiff as the photographer of the Photograph'" and that the "'removal of said [CMI] was made by Defendants intentionally, knowingly, and with the intent to induce, enable, facilitate, or conceal their infringement of Plaintiff's copyright in the Photograph.'"

Hirsch v. CBS Broad. Inc., No. 17-cv-1860, 2017 WL 3393845, at *8 (S.D.N.Y. Aug. 4, 2017) (quoting complaint); *see also Devocean Jewelry LLC v. Associated Newspapers Ltd*., No. 16-cv-2150, 2016 WL 6135662, at *2 (S.D.N.Y. Oct. 19, 2016) (denying motion to dismiss DMCA claims with "relatively sparse" allegations of scienter). The Intercept alleges far more than the plaintiffs in these other cases, who themselves alleged enough to survive dismissal.

First, as to intent, The Intercept alleges that "Defendants intentionally removed author, title, copyright notice, and terms of use information from Plaintiff's copyrighted works in creating ChatGPT training sets." Compl. ¶ 43. This itself is enough. *See Hirsch*, 2017 WL 3393845, at *8. But The Intercept alleges other facts too. As discussed, for example, it alleges that it published its works with specified CMI, while approximations of ChatGPT's training data contain copies of The Intercept's works without CMI. *See* Compl. ¶¶ 32, 38. If the ChatGPT training data lacks CMI, it is at least plausible that the data's creators—Defendants—intentionally removed it. Indeed, that is the ***only*** plausible explanation. Why else would the CMI be gone?

OpenAI disputes that The Intercept plausibly alleges intent largely by reprising Microsoft's flawed arguments regarding CMI removal. Like Microsoft, it ignores The Intercept's allegations about the public approximations of ChatGPT's training sets and focuses on the supposedly speculative inference from the absence of CMI in ChatGPT's outputs to the absence of CMI in the training data. Also like Microsoft, its argument fails on its own terms.

For one, OpenAI assumes that The Intercept can allege intent only if ChatGPT has regurgitated The Intercept's own works. *See* OpenAI Mot at 13. That assumption is unsupported and incorrect. The Intercept alleges that (1) ChatGPT's training sets include both The Intercept's and others' news articles, and (2) ChatGPT's outputs regurgitate copyright-protected works of journalism without CMI. *See* Compl. ¶¶ 4, 35. Absent any reason to believe that Defendants treat

different news organizations' differently—which the Complaint does not allege, and which in fact there is not—then there is one natural inference: if CMI-less regurgitations of some news articles results from the intentional removal of CMI, then Defendants remove all their CMI, including The Intercept's. Plus, a DMCA claim does not require any actual infringement of plaintiff's copyright. *See Murphy v. Millennium Radio Grp. LLC*, No. 08-cv-1743, 2015 WL 419884, at *5 (D.N.J. Jan. 30, 2015). Since Defendants know that regurgitation happens, The Intercept has sufficiently alleged that the requisite intent and knowledge existed at the time the CMI was removed, regardless of the extent to which regurgitation of The Intercept's works eventually occurs.

Like Microsoft, OpenAI also argues that it is too speculative to infer that the absence of CMI from ChatGPT outputs results from an intentional choice to remove CMI from the training data. *See* OpenAI Mot. at 13. Its argument fails for the same reason: plausibly, the outputs lack CMI because the inputs do, and The Intercept need not refute unspecified alternative possibilities.

OpenAI cites *Tremblay* v. *OpenAI, Inc.*, No. 23-cv-03223, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024), which found intent lacking, but the allegations there are distinguishable. For one, the plaintiffs included excerpts of ChatGPT outputs "that include multiple references to Plaintiffs' names," thus undermining the plaintiffs' own allegations that OpenAI had removed the CMI. *Id.* at *4. The Intercept's complaint contains no similar self-defeating allegations.[12] Further, while *Tremblay* held insufficient the bare allegation that defendants removed CMI "by design," *id.*, The

[12] OpenAI attempts to manufacture one by pointing to the allegation that outputs sometimes provide author and title. *See* OpenAI Mot. at 15 (citing Compl. ¶ 40). But it misleadingly fails to cite the rest of the paragraph: that outputs do so only "because other material used in a training set references the author or title in the text of such material (e.g., a Wikipedia article discussing the underlying works of journalism) ." Compl. ¶ 40. Thus, when an output references author or title, it is ***not*** because Defendants left the CMI in The Intercept's article. It is because the content of someone else's article happened to attribute the material to The Intercept.

Intercept has alleged more, including the absence of CMI in the approximated datasets. The Intercept has plausibly alleged that Defendants' removal of its CMI was intentional.

Second, The Intercept plausibly alleges that Defendants removed its CMI knowing, or with reason to know, "that it will induce, enable, facilitate, or conceal an infringement." 17 U.S.C. § 1202(b). The relevant infringement can be a third party's or the defendant's. *See Mango*, 970 F.3d at 172. If the infringement is a future one, it need not be certain; it only must be "likely." *Id*.

The Intercept has easily cleared this bar too. It alleges not only that Defendants had the requisite scienter—which by itself is enough under *Hirsch*—but gives further reasons why. For one, Defendants had reason to know that their CMI removal would conceal their own infringements: because any ChatGPT responses that incorporated or regurgitated The Intercept's works would lack CMI, those responses would not communicate to ChatGPT users that they infringed The Intercept's copyright. *See* Compl. ¶ 47. Defendants also had reason to know that removal would induce, enable, or facilitate infringement by ChatGPT users. That is because they promote ChatGPT as a tool that can be used to generate content for a future audience, and at least some users would be less likely to distribute ChatGPT's responses based on The Intercept's works if they knew those works were copyrighted. *See id*. ¶¶ 48-49. The future infringements are plainly likely, as the award-winning website, Copyleaks, found that "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." *Id*. ¶ 5. Further OpenAI has reached licensing deals with some media organizations and created tools "to allow copyright owners to block their work from being incorporated into training sets." *Id*. ¶¶ 62, 63. Viewed in the light most favorable to The Intercept, that evidences knowledge that the use in training, and output, of copyrighted-protected works constitutes infringement.

Defendants respond by attacking an argument The Intercept does not make: they say that because the training sets are secret, removing CMI from them does not conceal their own infringement from the public. *See* Microsoft Mot. at 21-22; OpenAI Mot. at 13-15. But that does not address The Intercept's actual scienter theory: Defendants knew or had reason to know that removing CMI during training results in CMI-less ***outputs*** that misinform the public about the source of the content Defendants' products provide to users. Defendants do argue that The Intercept has not specifically identified such outputs—ironic, given OpenAI's accusation that the New York Times "paid someone to hack OpenAI's products" by doing just that.[13] But such allegations are not necessary at this stage, particularly given the well-pleaded allegation regarding the extent to which ChatGPT's outputs plagiarize online content. *See* Compl. ¶ 4. That plagiarism makes it likely that ChatGPT's outputs infringe The Intercept's copyrights and that Defendants at least had reason to know that removing The Intercept's CMI would conceal their infringement. Moreover, The Intercept has plausibly alleged that Defendants had reason to know that their removal of CMI enables and facilitates their large-scale copying and use of copyright-protected material in their training sets by avoiding the practical problems for their products that would arise if Defendants had included CMI. *Id.* ¶¶ 34, 39, 50.

Microsoft also disputes that The Intercept has plausibly alleged that Defendants had reason to know that removal would lead ChatGPT users to infringe. It cites *Stevens v. Corelogic, Inc*., 899 F.3d 666 (9th Cir. 2018), which it says establishes a need to allege a "pattern of conduct" or "established modus operandi" showing that defendant was "aware of the probable future impact of its actions." *Id*. at 674. But *Stevens* was decided on summary judgment. And for that reason,

[13] Memorandum of Law in Support of OpenAI Defendants' Motion to Dismiss, 2, *The New York Times Company v. Microsoft Corp.*, No. 23-cv-11195 (S.D.N.Y. Feb. 26, 2024).

later courts in the Ninth Circuit have found its standard "inapposite" at the dismissal stage, as resolution of scienter issues "is more suited to summary judgment." *Izmo, Inc. v. Roadster, Inc.*, No. 18-cv-06092, 2019 WL 13210561, at *4 (N.D. Cal. Mar. 26, 2019); *see also Doe 1*, 672 F. Supp. 3d at 858 (holding that *Stevens*, as a summary judgment case, does not alter the rule that, at the pleading stage, "mental conditions generally need not be alleged with specificity"). Those holdings are consistent with this Circuit's standards for assessing scienter at the dismissal stage. Microsoft would impose too high a pleading bar.

Finally, Microsoft argues that The Intercept has not plausibly alleged "user behavior"—that users will be less likely to infringe The Intercept's copyrighted works if they know the works are copyrighted, either because they morally oppose stealing others' works or fear liability for doing so. Microsoft Mot. at 23; *see also* Compl. ¶ 49. But as Microsoft itself argues, one function of CMI is to "discourage piracy," *id*. at 10-11 (quoting S. Rep. 105-190 at 11 n.18 (1998)), which it does by alerting users that the work is copyrighted. Since Congress concluded that CMI will discourage infringement, that is surely enough to survive a motion to dismiss.

D. The Intercept states a claim under section 1202(b)(3).

A section 1202(b)(3) plaintiff must allege the following: "(1) the existence of CMI in connection with a copyrighted work; and (2) that a defendant distributed works or copies of works; (3) while knowing that CMI has been removed or altered without authority of the copyright owner or the law; and (4) while knowing, or having reasonable grounds to know that such distribution will induce, enable, facilitate, or conceal an infringement." *Mango*, 970 F.3d at 171 (cleaned up).

The Intercept has plausibly alleged each element. First, the existence of CMI on The Intercept's works is discussed above. *See* Section IV.C.1, *supra*.

Second, as to distribution, The Intercept alleges that Defendants "shared copies of Plaintiff's works from which author, title, copyright notice, and terms of use information has been

removed, with [each other] as part of Defendants' efforts to develop ChatGPT." Compl. ¶¶ 45-46. Defendants again argue that The Intercept has not identified the publicly available information on which it based its allegations (including that Microsoft, as opposed or in addition to OpenAI, was involved in the distribution). Microsoft Mot. at 18-19; OpenAI Mot. at 16. But that is false: it is the close working relationship between OpenAI and Microsoft. *See* Compl. ¶¶ 20-21. Equally off-base is OpenAI's criticism that The Intercept does not allege "when, why, or how" the distribution occurred. OpenAI Mot. at 16. That is plainly not required at the pleading stage. *See Pilla v. Gilat*, No. 19-cv-2255, 2020 WL 1309086, at *12 (S.D.N.Y. Mar. 19, 2020) ("Although Plaintiff does not allege how, when, or where this removal occurred, such details are not necessary at the pleading stage for a claim under the DMCA.").

OpenAI further argues on the second element that the Complaint does not allege distribution of "identical copies of Plaintiff's works," which it says—incorrectly—is necessary for a DMCA claim. OpenAI Mot. at 16. *See We Protesters, Inc. v. Sinyangwe*, No. 22-cv-9565, 2024 WL 1195417, at *9 (S.D.N.Y. Mar. 20, 2024) (holding that "close to identical" is enough). But the Complaint clearly does allege identicality. It alleges that Defendants "shared copies of Plaintiff's works" with each other. Compl. ¶¶ 45-46. Seeking to muddy the waters, OpenAI points to an allegation that Defendants later "adapted" some of the works they downloaded from the internet and argues that those adapted works cannot be identical. OpenAI Mot. at 16 (quoting Compl. ¶¶ 30-31). But that is a non-sequitur. The Complaint alleges that they shared "copies of Plaintiffs works," *i.e.*, the versions they downloaded from the internet but with CMI removed. Obviously a DMCA defendant cannot avoid liability on the self-fulfilling theory that once CMI is removed, the work is "adapted." Otherwise the provision would be a dead letter.

Third, The Intercept has alleged that Defendants distributed The Intercept's works knowing that CMI had been removed. As shown above, the Complaint alleges that Defendants themselves intentionally removed the CMI. So they must know the distributed copies lack CMI. Defendants do not appear to dispute this element beyond their incorrect arguments on intentional removal.

Fourth, as shown above, The Intercept alleges Defendants' knowledge or reason to know that distribution will induce, enable, facilitate, or conceal an infringement. OpenAI argues to the contrary because, supposedly, "Plaintiff does not identify a single instance of infringement of one of its works." OpenAI Mot. at 17. Not so. Defendants' copying thousands of The Intercept's copyrighted works infringes the reproduction right. *See* Compl. ¶¶ 31, 38; *Napster*, 239 F.3d at 1014 (holding that downloading of files containing copyrighted music violates the reproduction right). Indeed, OpenAI is aware of this, at least viewed in the light most favorable to The Intercept: it reached licensing deals with some large copyright owners and allows copyright owners to "block their work from being incorporated into training sets," Compl. ¶¶ 62, 63, suggesting it knows such incorporation infringes. And Defendants' distributing the works amongst themselves, with CMI removed, both conceals the sender's infringement and facilitates the recipient's by suggesting that the works are not copyrighted and thus may be lawfully copied or distributed further. Given the "lenient" pleading standards applicable to scienter, *Aaberg*, 2018 WL 1583037, at *9, this is more than enough to state a claim.

V. CONCLUSION

Defendants' motions to dismiss are based on unrecognized and unsupported legal theories, inflated pleading standards, and disregard for many of the Complaint's well-pleaded allegations. The Court should deny the motions and set the case for prompt trial. In the alternative, the Court should grant The Intercept leave to replead to correct any deficiencies the Court identifies. *See Shane Campbell Gallery, Inc. v. Frieze Events, Inc.*, 441 F. Supp. 3d 1, 4 (S.D.N.Y. 2020).

/s/ Matthew Topic

Jonathan Loevy (*pro hac vice*)
Michael Kanovitz (*pro hac vice*)
Lauren Carbajal (*pro hac vice*)
Stephen Stich Match (No. 5567854)
Matthew Topic (*pro hac vice*)

LOEVY & LOEVY
311 North Aberdeen, 3rd Floor
Chicago, IL 60607
312-243-5900 (p)
312-243-5902 (f)
jon@loevy.com
mike@loevy.com
carbajal@loevy.com
match@loevy.com
matt@loevy.com

Counsel for Plaintiff

Dated: May 6, 2024