

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

THE INTERCEPT MEDIA, INC.,

Plaintiff,

v.

No. 1:24-cv-01515-JSR

OPENAI, INC., OPENAI GP, LLC,  
OPENAI, LLC, OPENAI OPCO LLC,  
OPENAI GLOBAL LLC, OAI  
CORPORATION, LLC, OPENAI  
HOLDINGS, LLC, and MICROSOFT  
CORPORATION

Defendants.

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

**TABLE OF CONTENTS**

**I. INTRODUCTION.....1**

**II. ARGUMENT.....3**

**A. The Intercept has Article III standing .....3**

**1. The Intercept has alleged a concrete injury analogous to copyright infringement .....3**

**2. If standing requires dissemination, The Intercept is entitled to jurisdictional discovery .....9**

**B. The Intercept states a claim .....11**

**1. Defendants have abandoned many of their arguments for dismissal .....11**

**2. The Section 1202(b)(1) claim is not time-barred.....13**

**3. The Intercept plausibly alleged scienter under Section 1202(b)(1).....16**

**4. The Intercept plausibly alleges the Section 1202(b)(3) claims .....21**

**5. The Intercept plausibly alleges Microsoft’s involvement in the training process .....23**

**III. CONCLUSION .....24**

**TABLE OF AUTHORITIES**

	<i>Page</i>
 <i>Cases</i>	
<i>A&amp;M Recs., Inc. v. Napster, Inc.</i> , 239 F.3d 1004 (9th Cir. 2001) .....	3
<i>Aaberg v. Francesca’s Collections, Inc.</i> , No. 17-cv-115, 2018 WL 1583037 (S.D.N.Y. Mar. 27, 2018) .....	2, 16
<i>Abbas v. Dixon</i> , 480 F.3d 636 (2d Cir. 2007) .....	13
<i>ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.</i> , 667 F. Supp. 3d 411 (S.D. Tex. 2023) .....	23
<i>Amidax Trading Grp. v. S.W.I.F.T. SCRL</i> , 671 F.3d 140 (2d Cir. 2011) .....	9
<i>Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023) .....	4
<i>Devocean Jewelry LLC v. Associated Newspapers Ltd.</i> , No. 16-cv-2150, 2016 WL 6135662 (S.D.N.Y. Oct. 19, 2016) .....	16
<i>Dodds v. Cigna Sec., Inc.</i> , 12 F.3d 346 (2d Cir. 1993) .....	16
<i>Dowling v. U.S.</i> , 473 U.S. 207 (1985) .....	5
<i>Entick v. Carrington</i> , 2 Wils. K. B. 275, 291, 95 Eng. Rep. 807 (K. B. 1765) .....	7
<i>F. W. Woolworth Co. v. Contemp. Arts</i> , 344 U.S. 228 (1952) .....	6
<i>Feld v. Feld</i> , 783 F. Supp. 2d 76 (D.D.C. 2011) .....	7
<i>Gary Friedrich Enterprises, LLC v. Marvel Enterprises, Inc.</i> , 713 F. Supp. 2d 215 (S.D.N.Y. 2010) .....	5
<i>Grace v. Apple, Inc.</i> , 328 F.R.D. 320 (N.D. Cal. 2018) .....	6
<i>Hirsch v. CBS Broad. Inc.</i> , No. 17-cv-1860, 2017 WL 3393845 (S.D.N.Y. Aug. 4, 2017) .....	16
<i>Hirsch v. Rehs Galleries, Inc.</i> , No. 18-cv-11864, 2020 WL 917213 (S.D.N.Y. Feb. 26, 2020) .....	13, 16
<i>In re Flag Telecom Holdings, Ltd. Sec. Litig.</i> , 236 F.R.D. 177 (S.D.N.Y. 2006) .....	19
<i>Jewell-La Salle Realty Co. v. Buck</i> , 283 U.S. 202 (1931) .....	6
<i>Kamen v. Am. Tel. &amp; Tel. Co.</i> , 791 F.2d 1006 (2d Cir. 1986) .....	9

*Mango v. BuzzFeed, Inc.*, 970 F.3d 167 (2d Cir. 2020).....8, 17

*Mercer v. Jericho Hotels, LLC*, No. 19-cv-5604, 2019 WL 6117317 (S.D.N.Y. Nov. 18, 2019).....9

*Murphy v. Millennium Radio Grp. LLC*, No. 08-cv-1743, 2015 WL 419884 (D.N.J. Jan. 30, 2015) ..... 18

*Pilla v. Gilat*, No. 19-cv-2255, 2020 WL 1309086 (S.D.N.Y. Mar. 19, 2020).....22, 24

*Psihoyos v. John Wiley & Sons, Inc.*, 748 F.3d 120 (2d Cir. 2014).....14

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

*Shultz v. Congregation Shearith Israel of City of New York*, 867 F.3d 298 (2d Cir. 2017) .....22

*TransUnion LLC v. Ramirez*, 594 U.S. 413 (2021) .....3, 5, 6

*Tremblay v. OpenAI, Inc.*, No. 23-cv-03223, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024) .....18

*U.S. v. Spear*, No. 2:22-cv-00439, 2023 WL 6463039 (D. Idaho Oct. 4, 2023) .....7

*Warner Chappell Music, Inc. v. Nealy*, 144 S. Ct. 1135 (2024).....15

**Statutes**

17 U.S.C § 1202(c)(1).....4

17 U.S.C. § 106(2) .....4

17 U.S.C. § 106(3) .....5

17 U.S.C. § 106(a) .....3

17 U.S.C. § 106A(a) .....8

17 U.S.C. § 1202(b)(1) .....16

17 U.S.C. § 1202(b)(3) .....23

17 U.S.C. § 507(b) .....14

**Other Authorities**

Charlotte Tobitt, OpenAI content boss: ‘Incumbent’ on us to help small publishers, not just the giants, PressGazette (May 30, 2024) .....14

Fed. R. Civ. P. 8(c) .....13

Intelligencer Staff, Satya Nadella on Hiring the Most Powerful Man in AI, *Intelligencer*, (Nov. 21, 2023).....22

Lee Fang, Defense Tech Startup Founded by Trump’s Most Prominent Silicon Valley Supporters Wins Secretive Military AI Contract, *The Intercept* (Mar. 9, 2019).....15

Lee Fang, Google Continues Investment in Military and Police AI Technology Through Venture Capital Arm, *The Intercept* (July 23, 2019).....15

OpenAI and journalism (Jan. 8, 2024).....19

Pub. L. No. 101-650, 104 Stat. 5128 (1990).....8

Pub. L. No. 105-304, 122 Stat. 2860 (1998).....8

Restatement (Second) of Torts § 163 (1977).....6

Restatement (Second) of Torts, ¶ 577 cmt. b (1977).....5

Sam Biddle, Artificial Intelligence Experts Issue Urgent Warning Against Facial Scanning with a “Dangerous History,” *The Intercept* (Dec. 6, 2018) .....15

Sam Biddle, Facebook Uses Artificial Intelligence to Predict Your Future Actions for Advertisers, Says Confidential Document, *The Intercept* (Apr. 13, 2018).....15

Sam Biddle, The Internet’s New Favorite AI Proposes Torturing Iranians and Surveilling Mosques, *The Intercept* (Dec. 8, 2022).....16

Tom B. Brown et al, Language Models are Few-Shot Learners, 8 (July 22, 2020) .....14

U.S. Const. art. I § 8 cl. 8.....7

U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral Rights in the United States*, 90, 98 (April 2019).....8

## I. INTRODUCTION

The Court granted The Intercept leave to provide additional details. It has done that in spades. It listed all the works at issue. It showed that its articles are in Defendants' training sets. It described the processes by which Defendants intentionally removed CMI from its news stories. It provided more details about Microsoft's involvement. It added allegations about Defendants' knowledge that its conduct conceals and abets infringement. And it shows that ChatGPT is capable of regurgitating The Intercept's copyright-protected works. As a result, Defendants have abandoned many of their arguments for dismissal. Their remaining arguments do not warrant it.

On standing, the Court should deny Defendants' motion for reasons already addressed: The Intercept has alleged a concrete injury based on Defendants' interference with its copyright-protected property. Its injuries are analogous to copyright infringement; indeed, its injuries might well *be* copyright infringement—a point made through new allegations in the FAC that Defendants ignore. Defendants try once more to analogize the DMCA to defamation and attribution and argue that these injuries require dissemination. But copyright infringement, the relevant historical analogy, has never required dissemination—a point Defendants have never disputed.

If the Court concludes Article III requires dissemination, though, it should at least order jurisdictional discovery. After the motion hearing, The Intercept did exactly what OpenAI's counsel said it should do: “sweet talk ChatGPT into” producing regurgitations of The Intercept's works. MTD Hearing Tr. at 11:12. Now that The Intercept has shown that ChatGPT can regurgitate its works, Defendants take the exact opposite position: that The Intercept's “sweet talking” is “obviously not sufficient to establish an Article III injury” because “The Intercept itself has elicited these outputs.” OpenAI Supp. Br. at 14, ECF No. 89. But what *other* users have done is a fact uniquely within Defendants' possession. FAC ¶ 63, ECF No. 87. And while The Intercept only produced regurgitations based on artificial prompting, that is because OpenAI has been hard

at work stopping ChatGPT's regurgitations for purposes of avoiding further copyright liability. FAC ¶ 66. The Intercept has therefore laid sufficient foundation to discover what past users did years ago, when Defendants were far more cavalier about copyright. If the Court adopts Defendants' theory of standing as requiring dissemination, despite standing and copyright case law to the contrary, it should order discovery into whether it is satisfied.

Defendants' merits arguments are simply fact disputes that do not warrant dismissal at this early stage. The OpenAI Defendants seek dismissal of Count I under the Copyright Act's three-year limitations period. They argue that the specific instances of CMI removal alleged in the FAC occurred more than three years before The Intercept filed suit—seeking to leverage their more recent secrecy about the contents of their training sets. But The Intercept has plausibly alleged that OpenAI employed the same CMI removal processes within the last three years. And fact questions remain as to whether the discovery rule tolls that period further.

Defendants next seek dismissal on scienter grounds. But their arguments again ignore this Circuit's "lenient" pleading rules and many of the FAC's new allegations about their actual and constructive knowledge. *Aaberg v. Francesca's Collections, Inc.*, No. 17-cv-115, 2018 WL 1583037, at \*9 (S.D.N.Y. Mar. 27, 2018). These allegations support the plausible inference that Defendants both knew and had reason to know that their CMI removal would conceal their infringement from ChatGPT users, and further their users' infringement.

Finally, Defendants seek dismissal of the Section 1202(b)(3) claims, and the claims related to Microsoft's involvement in OpenAI's training. But the FAC's allegations—including an admission by Microsoft's CEO that "we have the data,"—plausibly support that Microsoft does, in fact, have the data, used it to train AI models, and distributed it to and from OpenAI.

The Court should deny the motions to dismiss.

## II. ARGUMENT

### A. The Intercept has Article III standing.

#### 1. The Intercept has alleged a concrete injury analogous to copyright infringement.

A plaintiff satisfies Article III’s concreteness requirement if it has suffered either a historical or common-law injury, or an “analogue” to such an injury, which need not be an “exact duplicate.” *TransUnion LLC v. Ramirez*, 594 U.S. 413, 424 (2021).<sup>1</sup> The Intercept has identified copyright infringement as that historical analogue. Though not an exact duplicate of the DMCA injuries The Intercept suffered, it is close enough under Second Circuit and Supreme Court precedent to satisfy Article III. That is largely because The Intercept’s DMCA injuries likely *are* copyright infringement, or at least very close.

Start with the Section 1202(b)(1) claim for unlawful removal of CMI. As the FAC alleges, Defendants removed The Intercept’s CMI through a technical process that involved downloading The Intercept’s copyright-protected works and applying computer algorithms to create further copies with CMI deliberately removed. *See* FAC ¶¶ 45-53. Their removing CMI in this way—through creating unauthorized copies of The Intercept’s articles—*prima facie* violates The Intercept’s exclusive right “to reproduce the copyrighted work.” 17 U.S.C. § 106(a); *see also A&M Recs., Inc. v. Napster, Inc.*, 239 F.3d 1004, 1014 (9th Cir. 2001) (holding that downloading files violates the reproduction right). Defendants cannot reasonably dispute that such copying is *prima facie* copyright infringement. Indeed, they did not dispute it when other news organizations sued them for infringing their copyright, electing instead to stake their defense on fair use.<sup>2</sup> So

---

<sup>1</sup> The OpenAI Defendants initially argued that The Intercept failed Article III’s particularity requirement. *See* OpenAI Mot. at 6, ECF No. 53; OpenAI Reply at 1-2, ECF No. 80. They appear to have abandoned that argument, as it does not appear in their supplemental brief.

<sup>2</sup> *See* Memorandum of Law in Support of OpenAI Defendants’ Motion to Dismiss, 3, *The New York Times Company v. Microsoft Corporation*, No. 23-cv-11195 (S.D.N.Y. Feb. 26, 2024), ECF



because The Intercept’s Section 1202(b)(1) injuries are *actual prima facie* copyright infringement, these injuries are at least *analogous* to copyright infringement under Article III.<sup>3</sup>

The CMI removal Defendants inflicted on The Intercept is also analogous to a violation of the right to “prepare derivative works.” 17 U.S.C. § 106(2). *See* P. Opp. at 6, ECF No. 78. That right allows the copyright owner to alter the work’s “meaning” or “provide new information.” *Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 541 (2023). And at least some CMI does just that. Common experience shows that a headline—a form of CMI (*i.e.*, a title)—can substantially impact the ensuing article’s meaning. 17 U.S.C § 1202(c)(1). For instance, one of The Intercept’s most well-known articles is called “The Drone Papers,” an allusion to the Pentagon Papers that thus evokes a patriotic act of whistleblowing, done at great personal risk, to shed light on egregious government misconduct. *See* FAC Ex. 1 at 337 (listing The Drone Papers as a work at issue in this case). The Intercept’s article, which exposes secrets involving the U.S. government’s drone program, would lack that evocation without its headline. So the removal of that headline, like the others in this case, interferes with The Intercept’s exclusive right to define the meaning of its articles in much the same as the work to prepare derivative works.<sup>4</sup>

---

No. 52; Defendant Microsoft Corporation’s Memorandum in Support of Partial Motion to Dismiss the Complaint, 2, *The New York Times Company v. Microsoft Corporation*, No. 23-cv-11195 (S.D.N.Y. Mar. 4, 2024), ECF No. 65; Defendant Microsoft Corporation’s Memorandum in Support of Partial Motion to Dismiss the Complaint, 1, *Daily News, LP v. Microsoft Corporation*, No. 24-cv-03285 (S.D.N.Y. June 11, 2024), ECF No. 77; Memorandum of Law in Support of OpenAI Defendants’ Motion to Dismiss, 2, *Daily News, LP v. Microsoft Corporation*, No. 24-cv-03285 (S.D.N.Y. June 11, 2024), ECF No. 82.

<sup>3</sup> As The Intercept has explained, it did not bring copyright infringement claims because such claims require registration, and, unlike print media, there is no bulk registration option for online-only news articles. *See* MTD Hearing Tr. at 8:3-11.

<sup>4</sup> The same goes for other forms of CMI. For instance, a copyright notice attributing an article to The Intercept or one of its authors conveys a sense of “fearless journalism” and trustworthiness that would not be conveyed without that information. FAC ¶ 8.

Take next the Section 1202(b)(3) claims for distributing works without CMI. Violations of that section will often infringe the right to “distribute” copies of the work. 17 U.S.C. § 106(3). By definition, both require the distribution. While the distribution between OpenAI and Microsoft might not qualify as distribution “to the public,” *see id.*, Article III standing does not require an “exact duplicate” of this historical injury. *TransUnion*, 594 U.S. at 424. Indeed, *TransUnion* supports the proposition that distribution to just one other person *can* satisfy Article III. There, the Court divided the plaintiffs into two groups based on whether they met the “publication” element of common law defamation. *Id.* at 432. Those who met it had standing; those who failed it did not. And, crucially, defamation’s publication element is satisfied by communication “to a single individual other than the one defamed.” Restatement (Second) of Torts, ¶ 577 cmt. b (1977).<sup>5</sup> The Intercept has alleged just that.

These analogies are far closer than that between (1) encumbering voting shares through a fund’s amendment to its bylaws and (2) common law trespass to chattels, which the Second Circuit found sufficient in *Saba Capital Cef Opportunities 1, Ltd. v. Nuveen Floating Rate Income Fund*, 88 F.4th 103 (2d Cir. 2023). Unlike the claims here, it is doubtful the injury in *Nuveen* could even *theoretically* satisfy the common law tort. Trespass to chattels requires the defendant to assume “physical control” of the chattel, which does not occur when a fund amends its bylaws. *Gary Friedrich Enterprises, LLC v. Marvel Enterprises, Inc.*, 713 F. Supp. 2d 215, 231 (S.D.N.Y. 2010) (quoting *Dowling v. U.S.*, 473 U.S. 207, 217 (1985)). It was instead enough that the bylaw amendment caused something “analogous to a property-based injury.” *Nuveen*, 88 F.4th at 116.

---

<sup>5</sup> Microsoft guesses that *TransUnion* would have rejected standing for a plaintiff who alleged distribution from one credit reporting agency to another. *See* Microsoft Supp. Br. at 4. *TransUnion* did not address that question: it concerned only credit files that were distributed either to potential creditors or not outside *TransUnion*. *See TransUnion*, 594 U.S. at 433. And given the Court’s reliance on the publication element, Microsoft is wrong about the result it would have reached.

Here, The Intercept’s DMCA injuries might well *be* copyright infringement. The copyright infringement analogy is close, and certainly close enough.

Having analogized their injuries to copyright infringement, The Intercept need do no more. That is because copyright infringement, without more, has always been actionable, including when its results are otherwise “uninjurious and unprofitable,” *F. W. Woolworth Co. v. Contemp. Arts*, 344 U.S. 228, 233 (1952), and even when, other than the infringement itself, there is “no showing as to actual loss,” *Jewell-La Salle Realty Co. v. Buck*, 283 U.S. 202, 208 (1931). *See also* P. Opp. at 7-8 (citing additional authorities). In this respect copyright infringement is like quintessential common-law injuries, such as trespass to land and trespass to chattels, that recognize a distinct and concrete harm from interference with property. *See* Restatement (Second) of Torts § 163 (1977) (providing that trespass to land is actionable without any additional showing of harm); *Grace v. Apple, Inc.*, 328 F.R.D. 320, 346 (N.D. Cal. 2018) (listing states where, “interference is enough” for liability for trespass to chattels, even without a further “showing of harm”). The Intercept suffered the same kind of harm when Defendants interfered with their copyright-protected articles by removing CMI from them and distributing them to each other without authority.<sup>6</sup>

Defendants resist this conclusion on three main grounds. First, they accuse The Intercept of advancing a “copyright exception” to the concreteness requirement. OpenAI Supp. Br. at 11. Not so. Under *TransUnion*, courts *determine* concreteness by considering whether the asserted injury was, or is analogous to, one “traditionally recognized as providing a basis for a lawsuit in American courts.” *TransUnion*, 594 U.S. at 424 (cleaned up). Such injuries can be “intangible

---

<sup>6</sup> Were there any doubt, these considerations dispense with Microsoft’s accusation that the Intercept has “begg[ed] the question of whether CMI removal is indeed analogous to the sorts of invasions of copyright interests cognizable historically or at common law.” Microsoft Supp. Br. at 4. The Intercept has not simply assumed that the harms it suffered are analogous to copyright infringement, but affirmatively shown it.

harms,” including those “specified by the Constitution itself.” *Id.* at 425. The Constitution, of course, specifies copyright. U.S. Const. art. I § 8 cl. 8. And as The Intercept has explained, an unbroken line of Congresses, from the Founding to today, has understood the Copyright Clause to authorize suits based solely on infringement, without any further showing. *See* P. Opp. at 7-8 (citing historical sources). Like all other Article III injuries, this historical tradition is what *makes* copyright infringement a concrete injury. The Intercept’s DMCA injuries, as analogous to copyright infringement, are concrete for the same reason.

Further, Defendants’ approach of second-guessing historical tradition would upend standing law and contradict *TransUnion*’s central holding. If interference with property is no longer concrete, then it hard to see how perhaps *the* quintessential common-law tort—pure trespass to land—could even be heard in federal court, despite its pre-Founding provenance. *See Entick v. Carrington*, 2 Wils. K. B. 275, 291, 95 Eng. Rep. 807, 817 (K. B. 1765).<sup>7</sup> Such a result is irreconcilable with *TransUnion*, under which the concreteness requirement derives its content from just these historical claims. This argument is meritless.

Second, Defendants argue that CMI is not property and that the DMCA does not create property rights. Microsoft Supp. Br. at 4, ECF No. 88; OpenAI Supp. Br. at 12, ECF No. 89. This misses the point. The point is that Defendants have interfered with copies of its copyright-protected *works*—the news articles—by removing CMI from them and distributing them without consent.

---

<sup>7</sup> Though such cases would not arise under federal question or diversity jurisdiction (for failure to satisfy the amount-in-controversy requirement), they could be heard in federal court if, for example, the United States is a party, *see, e.g., U.S. v. Spear*, No. 2:22-cv-00439, 2023 WL 6463039, \*6-7 (D. Idaho Oct. 4, 2023) (trespass action brought as United States), or through the Court’s exercise of supplemental jurisdiction, *see, e.g., Feld v. Feld*, 783 F. Supp. 2d 76, 77 (D.D.C. 2011) (trespass action brought as counterclaim).

Third, Defendants resist the analogy to copyright infringement because, they say, (1) any injury under the DMCA must be based on attribution, and (2) attribution lacks the required historical pedigree. *See* Microsoft Supp. Br. at 4-5. But the first premise is flawed for numerous reasons. It is contradicted by Microsoft’s own authority which correctly notes that, unlike any genuine attribution statute, “nothing in section 1202 affirmatively requires that an author be credited,” and recommends that Congress enact a new statute that would actually protect an author’s interest in attribution. U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral Rights in the United States*, 90, 98 (April 2019). Section 1202 prohibits removing CMI. But it does not require adding CMI. So one can comply with it without attributing anything to anyone.

Defendants’ argument is also inconsistent with statute. If attribution were the purpose of Section 1202(b), a statute prohibiting distribution of CMI-less works—Section 1202(b)(3)—would have been entirely sufficient, and a removal-based statute—Section 1202(b)(1)—would have been superfluous. So attribution is not the DMCA’s purpose. The purpose is “to provide broad protections to copyright owners.” *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 n.2 (2d Cir. 2020). And, of course, Congress knows how to grant actual attribution rights when it wants to. It did that when it enacted the Visual Artists Rights Act (“VARA”) just eight years before the DMCA. *See* Pub. L. No. 101-650, 104 Stat. 5128 (1990) (enacting VARA); Pub. L. No. 105-304, 112 Stat. 2860 (1998) (enacting DMCA). VARA expressly grants certain visual artists “rights of attribution and integrity,” including the right “to claim authorship of that work.” 17 U.S.C. § 106A(a). The DMCA contains nothing similar. So it is irrelevant that attribution might lack sufficient historical pedigree. It is simply not analogous to the DMCA claims at issue here. The right analogy is copyright infringement. And under that analogy, The Intercept has standing.

**2. If standing requires dissemination, The Intercept is entitled to jurisdictional discovery.**

Defendants' theory of standing would require The Intercept to allege that its CMI-less works have been disseminated to the public. *See* Microsoft Supp. Br. at 3; OpenAI Supp. Br. at 13-16. That theory is wrong for reasons already given: copyright infringement, the relevant analogy, does not require dissemination. *See* Section II.A.1, *supra*; P. Opp. at 8. But if the Court concludes otherwise, it should order jurisdictional discovery into the extent to which ChatGPT and Microsoft Copilot have disseminated The Intercept's works.

Discovery is often an appropriate device for assessing subject matter jurisdiction. *See Amidax Trading Grp. v. S.W.I.F.T. SCRL*, 671 F.3d 140, 149 (2d Cir. 2011) (holding that "a court should take care to give the plaintiff ample opportunity to secure and present evidence relevant to the existence of jurisdiction"). It is especially appropriate "where the facts are peculiarly within the knowledge of the opposing party." *Kamen v. Am. Tel. & Tel. Co.*, 791 F.2d 1006, 1011 (2d Cir. 1986). *Compare, e.g., Amidax*, 671 F.3d at 149 (affirming denial of jurisdictional discovery because the plaintiff, "not the defendants, is in control of the relevant jurisdictional evidence") with *Mercer v. Jericho Hotels, LLC*, No. 19-cv-5604, 2019 WL 6117317, at \*3 (S.D.N.Y. Nov. 18, 2019) (granting jurisdictional discovery on voluntary cessation exception to mootness because only the defendant knew its "substantive plans" for the future).

Here, Defendants are in sole possession of the relevant jurisdictional facts, at least on their theory of standing. On that theory, a user other than The Intercept needs to have prompted ChatGPT or Copilot to regurgitate The Intercept's articles—though one could be forgiven for thinking otherwise given OpenAI's counsel's argument that The Intercept itself should have "tried to sweet talk ChatGPT" into providing regurgitations. MTD Hearing Tr. at 11:12. The Intercept has no feasible way to learn what other users have done. But Defendants do. As the FAC alleges,

OpenAI has a repository of ChatGPT users' chat histories, including any regurgitations of The Intercept's articles, except when users have affirmatively disabled that feature. FAC ¶ 63. Jurisdictional discovery is warranted for exactly these informational asymmetries.

Further supporting jurisdictional discovery, The Intercept has laid the foundation for ChatGPT's capacity to disseminate its articles. That, and nothing else, was the purpose of the partial regurgitations attached to the FAC. *See* FAC Ex. 6.<sup>8</sup> True, the prompts that induced these regurgitations were somewhat artificial. But, as The Intercept plausibly alleges, users in the years predating the Complaint likely could have obtained regurgitations using more realistic prompts. And that is because, since the Complaint was filed, Defendants have been hard at work modifying their AI models not to regurgitate news stories. Indeed, as the FAC alleges, The Intercept tried to obtain the same regurgitations as the news publisher plaintiffs in a case filed less than two months before the FAC, *Daily News, LP v. Microsoft Corporation*, No. 24-cv-03285 (S.D.N.Y. Apr. 30, 2024). FAC ¶ 66. Even though The Intercept used the same methodology as those Plaintiffs, it failed. And in one case, ChatGPT responded, "I'm sorry, but I can't generate the original ending for the article or any copyrighted content." FAC ¶ 66. So, in just the last two months, ChatGPT has become far more resistant to regurgitation. It was therefore presumably far *less* resistant at the start of the limitations period more than three years ago. *See* Section II.B.2, *infra* (discussing limitations period). So because partial regurgitation is possible even now, jurisdictional discovery is plainly warranted to assess ChatGPT's regurgitations in years past, at least if standing requires dissemination—which, again, it does not. *See* Section II.A.1, *supra*.<sup>9</sup>

---

<sup>8</sup> Defendants spill much ink arguing that The Intercept's regurgitations do not in and of themselves establish an Article III harm. *See* Microsoft Supp. Br. at 3; OpenAI Supp. Br. at 13-16. Because The Intercept does not contend otherwise, it need not respond to these arguments.

<sup>9</sup> The Intercept did not raise jurisdictional discovery earlier because it was irrelevant to its primary theory of standing, on which it defended the Complaint at the motions hearing. It is appropriate

**B. The Intercept states a claim.**

**1. Defendants have abandoned many of their arguments for dismissal.**

While Defendants maintain that the FAC should be dismissed, their barrage of objections to the original Complaint is largely absent from their supplemental briefs. That is because the FAC has addressed many of the grounds for dismissal Defendants advanced against the original Complaint.<sup>10</sup> The abandoned arguments include:

*Identifying the works at issue.* Microsoft Mot. at 15, ECF No. 50; OpenAI Mot. at 10-12, ECF No. 53. The Intercept has now identified every work it owns that was published on its website. FAC ¶¶ 31-32; FAC Ex. 1. It has also identified every one of its 5,026 works present in a recreated approximation of OpenAI’s training set, WebText, and explained in more detail who created that approximation, and how. FAC ¶¶ 40-43; FAC Ex. 2. It has further alleged that Defendants are able to identify each one of The Intercept’s websites present in each of their training sets, giving Defendants further notice of the claims against them. FAC ¶ 42. Defendants’ supplemental briefs do not take issue with the adequacy of these allegations.

*Specifying which CMI was removed from which articles.* See Microsoft Mot. at 17-18. The FAC alleges that each type of CMI—author, title, copyright notice, and terms of use—was removed from each article. FAC ¶ 48. The copyright notice and terms of use are the same across all articles. *Id.* And The Intercept has identified the title and author of each URL that was, or potentially has been, included in OpenAI’s training sets. See FAC Ex. 1 (identifying author, title, and URL of each work); FAC Ex. 2 (identify URLs of all articles contained in recreated

---

now given recent developments, including the addition of new allegations regarding ChatGPT’s evolving incapacity to regurgitate and OpenAI’s possession of a database of ChatGPT prompts and responses. See FAC ¶¶ 63, 66.

<sup>10</sup> The Intercept does not concede that any of these grounds were meritorious to begin with, but addressed them out of an abundance of caution.



approximation of OpenAI's WebText training set). Microsoft's supplemental brief does not press this challenge.

*CMI removal from The Intercept's works.* See Microsoft Mot. at 15-16. The FAC provides substantially more detail about how Defendants removed CMI. It identifies the computer algorithms that OpenAI used when assembling WebText, explains that those algorithms are generally incapable of extracting CMI,<sup>11</sup> and provides examples of applications of those algorithms to The Intercept's articles. FAC ¶¶ 45-52. Those examples lack author, title, copyright notice, and terms of use information. FAC Exs. 3, 4. In addition, the FAC gave an expanded explanation of Defendants' use of Common Crawl data, a filtered version of which OpenAI admitted to using when training GPT-2. FAC ¶ 56. A similar filtered version, made publicly available by a nonprofit established by Microsoft's co-founder, contains content from 2,753 distinct URLs from The Intercept webpage, many of which are reproduced identically and lack any form of CMI. FAC ¶¶ 57-59; FAC Ex. 5. The FAC plainly alleges CMI removal from The Intercept's works, and Defendants' supplemental briefs say nothing to the contrary.<sup>12</sup>

*Intent to remove CMI.* See OpenAI Mot. at 13. The FAC alleges Defendants' intent to remove CMI. In particular, OpenAI has publicly stated that it used the Dagnet and Newspaper algorithms to extract text from websites, and these algorithms were expressly designed *not* to extract CMI. FAC ¶¶ 45-47. Because OpenAI employs highly skilled data scientists who would have known how the algorithms work, the FAC plausibly alleges Defendants' intent. FAC ¶ 50.

---

<sup>11</sup> One of the two algorithms, Newspaper, can be configured to extract title and author, but not copyright notice and terms of use. But, as the FAC alleges, OpenAI employed two algorithms to prevent redundancies, and the other algorithm, Dagnet, is incapable of extracting title and author. To ensure consistency across its text extraction, OpenAI plausibly configured Newspaper not to extract title and author. FAC ¶¶ 46-47.

<sup>12</sup> Microsoft continues to dispute its involvement in CMI removal. That point is addressed in Section II.B.5, *infra*.

Defendants do not dispute the adequacy of these allegations. *See* OpenAI Supp. Br. at 19 (outlining double scienter elements but not challenging intent).

**2. The Section 1202(b)(1) claim is not time-barred.**

For the first time in their supplemental brief, the OpenAI Defendants argue that the Section 1202(b)(1) claim against them is time-barred. In support, they point out that the FAC alleged that OpenAI assembled its WebText dataset, on which The Intercept relies, more than three years before The Intercept filed its Complaint. OpenAI Supp. Br. at 16-17; 17 U.S.C. § 507(b) (three-year statute of limitations). Microsoft wisely makes no such argument because it lacks any merit.

Statute of limitations is an affirmative defense. *See* Fed. R. Civ. P. 8(c). It is black letter law that litigants need not “affirmatively plead facts in avoidance of such defenses.” *Abbas v. Dixon*, 480 F.3d 636, 640 (2d Cir. 2007).<sup>13</sup> Dismissal on limitations grounds is thus reserved for cases where failure to comply with the statute of limitations is “clear from the face of the complaint, and matters of which the court may take judicial notice.” *Hirsch v. Rehs Galleries, Inc.*, No. 18-cv-11864, 2020 WL 917213, at \*3 (S.D.N.Y. Feb. 26, 2020). Thus, it is unwarranted when even “some doubt” remains. *Id.* Three reasons mandate rejecting this ground for dismissal.

First, the FAC plausibly alleges that OpenAI has continued to build training sets using the same methods that it did in the past. OpenAI has been entirely secret about how it trained its more recent GPT models. FAC ¶ 37. *See* MTD Hearing Tr. at 10:8-10 (OpenAI’s counsel answering “That is correct” to the Court’s question, “So am I right that your training sets are secret?”). So it is at least plausible at the pleading stage that its recent training used the same or similar methods as its past training. Thus, as the FAC alleges, OpenAI has continued to use text extraction methods

---

<sup>13</sup> For this reason, it is irrelevant whether or not The Intercept “made no attempt to supplement its pleadings” to address statute-of-limitations arguments the OpenAI Defendants made in other cases. OpenAI Supp. Br. at 18. This argument is also wrong for reasons explained below.

designed to remove CMI from news articles—methods OpenAI has never publicly disclaimed to have used more recently. FAC ¶ 54. Because OpenAI used thousands of The Intercept’s articles in its earlier training sets, it is also plausible at the pleading to infer that it did the same in its later ones. And contrary to OpenAI’s suggestion, The Intercept’s allegations are not simply about the works included in the older WebText dataset. Exhibit 1 to the FAC identifies the works at issue, and those works are dated up to June 12, 2024. FAC Ex. 1.

Second, nothing on the face of the FAC rules out that OpenAI removed CMI from the articles in the WebText training set in the three years before The Intercept filed its Complaint. Indeed, the FAC, and the articles it “incorporates by reference,” Microsoft Supp. Br. at 6, imply the opposite. OpenAI’s models are “trained from scratch,” according to a senior OpenAI executive. *See* 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/>; *see also* FAC ¶ 79 n.14 (incorporating article). And they are trained from scratch by “expand[ing]” past training sets, including WebText. FAC ¶ 40 (alleging that a later set, WebText2, is an “expanded version of WebText”).<sup>14</sup> Given its past practice of expanding WebText by starting from scratch, and the absence of any contrary information, it is plausible at the pleading stage that OpenAI continued this practice within the three-year limitations period, and in that period again removed CMI from The Intercept’s articles included in WebText.

Third, the face of the FAC does not preclude the discovery rule, which applies to suits under the Copyright Act. *See Psihoyos v. John Wiley & Sons, Inc.*, 748 F.3d 120, 124 (2d Cir. 2014) (construing 17 U.S.C. § 507(b), which applies to all claims under the Copyright Act,

---

<sup>14</sup> *See also* Tom B. Brown et al, Language Models are Few-Shot Learners, 8 (July 22, 2020), <https://arxiv.org/pdf/2005.14165>; FAC ¶ 56 n.10 (incorporating article).

including the DMCA, to embody the discovery rule).<sup>15</sup> That rule provides that a claim does not accrue until the plaintiff “discovers, or with due diligence should have discovered,” it. *Id.*

The OpenAI Defendants do not argue that The Intercept actually discovered its claim before February 28, 2021, three years before it filed the Complaint. Nothing in the FAC suggests it did. Instead, they argue that The Intercept was supposedly on “inquiry notice” because it published articles on “developments in artificial intelligence in 2018 and 2019.” OpenAI Mot. at 17 (citing articles listed in FAC Ex. 1 at 152, 169, 180, 217) . But none of those articles have anything to do with OpenAI, ChatGPT, large language models, or copyright. In the first two, The Intercept reports on little-known startups winning law enforcement and military contracts for using video-based AI to improve drone surveillance and targeting. In the third, The Intercept discusses the use of AI in facial recognition technology. And the fourth concern’s Facebook’s use of AI to predict its users’ behavior for purposes of selling targeted ads.<sup>16</sup> That these articles involve the broad concept of “artificial intelligence” is obviously not enough to put The Intercept “on inquiry” that OpenAI downloaded its articles and removed CMI from them. The OpenAI Defendants also fail to mention that OpenAI, ChatGPT, large language models, or copyright were not reported on

---

<sup>15</sup> The OpenAI Defendants cite a dissenting Supreme Court opinion opining that the Copyright Act does not embody the discovery rule. *See* OpenAI Supp. Br. at 17. But the full Court did not reach the question, *see Warner Chappell Music, Inc. v. Nealy*, 144 S. Ct. 1135, 1137 (2024), and a nonbinding dissent obviously does not overrule binding Circuit precedent.

<sup>16</sup> *See* FAC Ex. 1 at 152, 169, 180, 217 (citing Lee Fang, Google Continues Investment in Military and Police AI Technology Through Venture Capital Arm, *The Intercept* (July 23, 2019), <https://theintercept.com/2019/07/23/google-ai-gradient-ventures/>; Lee Fang, Defense Tech Startup Founded by Trump’s Most Prominent Silicon Valley Supporters Wins Secretive Military AI Contract, *The Intercept* (Mar. 9, 2019); Sam Biddle, Artificial Intelligence Experts Issue Urgent Warning Against Facial Scanning with a “Dangerous History,” *The Intercept* (Dec. 6, 2018), <https://theintercept.com/2018/12/06/artificial-intelligence-experts-issue-urgent-warning-against-facial-scanning-with-a-dangerous-history/>; Sam Biddle, Facebook Uses Artificial Intelligence to Predict Your Future Actions for Advertisers, Says Confidential Document, *The Intercept* (Apr. 13, 2018), <https://theintercept.com/2018/04/13/facebook-advertising-data-artificial-intelligence-ai/>).

by The Intercept until December 2022, barely a year before The Intercept filed its Complaint. *See* FAC Ex. 1 at 38 (citing Sam Biddle, The Internet’s New Favorite AI Proposes Torturing Iranians and Surveilling Mosques, *The Intercept* (Dec. 8, 2022), <https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/>).

The OpenAI Defendants also argue that The Intercept was put on inquiry by the articles, blog posts, and other webpages cited in the FAC and published in 2020 or earlier. *See* OpenAI Supp. Br. at 17-18. But according to their own case, a plaintiff is not required to conduct *any* inquiry until it is presented with a “probability” of a violation. *Dodds v. Cigna Sec., Inc.*, 12 F.3d 346, 350 (2d Cir. 1993). This principle applies with full force under the Copyright Act’s discovery rule, where the copyright owner has no “general duty to police the internet” for possible violations. *Hirsch*, 2020 WL 917213, at \*5. Whether or not The Intercept’s exposure to these obscure sources would have put The Intercept on inquiry, nothing in the FAC or OpenAI’s supplemental brief remotely suggests The Intercept was aware of them more than three years before it sued. The Court should reject this baseless defense.

### **3. The Intercept plausibly alleged scienter under Section 1202(b)(1).**

Equally baseless is Defendants’ argument on the second scienter element: that they knew, or had reasonable grounds to know, that their CMI removal was likely to “induce, enable, facilitate, or conceal an infringement.” 17 U.S.C. § 1202(b)(1). The Intercept has easily cleared the low bar for pleading scienter, a bar that applies in DMCA cases. *See Aaberg v. Francesca’s Collections, Inc.*, No. 17-cv-115, 2018 WL 1583037, at \*9 (S.D.N.Y. Mar. 27, 2018); *Hirsch v. CBS Broad. Inc.*, No. 17-cv-1860, 2017 WL 3393845, at \*8 (S.D.N.Y. Aug. 4, 2017); *Devocean Jewelry LLC v. Associated Newspapers Ltd.*, No. 16-cv-2150, 2016 WL 6135662, at \*2 (S.D.N.Y. Oct. 19, 2016). The FAC plausibly alleges scienter on three grounds.

*Concealment from users of infringement in training.* Defendants infringed The Intercept’s copyright by copying its articles into their training sets without permission. FAC ¶ 61. Defendants plausibly know this infringed because (1) OpenAI has recently allowed publishers to prevent OpenAI from using their articles during training, plausibly to avoid copyright liability, and (2) OpenAI has made deals with several publishers (deals never offered The Intercept) that granted OpenAI the right to use their articles in training, plausibly because it recognized that training is not fair use. FAC ¶¶ 77-80.

Defendants also had reason to know that removing CMI during training would conceal that infringement from users. A ChatGPT response derived from an Intercept article that *contained* The Intercept’s CMI would inform the user that the response blossomed out of an infringing copy in the training set. After all, “ChatGPT responses are the product of its training sets.” FAC ¶ 82. Thus, a response that *omits* CMI conceals that infringement. And, critically, ChatGPT omits CMI precisely because Defendants removed the CMI during training: if ChatGPT were trained on CMI, it would have produced CMI. FAC ¶ 70.

Further, Defendants had reason to know that this phenomenon would likely occur in practice. Nearly 60% of ChatGPT’s responses contain material identical to preexisting text, and ChatGPT’s training sets contain thousands of The Intercept’s copyright-protected works—in fact, The Intercept owns one of the top 1,000 web domains present in WebText. FAC ¶¶ 5, 41. ChatGPT is a serial plagiarist and had thousands of The Intercept’s works to plagiarize. So Defendants had reason to know it would produce responses derived from The Intercept’s copyright-protected articles, which concealed that they blossomed out of infringing copies.<sup>17</sup> Put

---

<sup>17</sup> Of course, whether ChatGPT later produced any concealing responses is irrelevant to whether Defendants had “reasonable grounds to know” it would. See *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020) (holding that plaintiffs may establish scienter by alleging “likely future

simply, they had reason to know that removing CMI during training would conceal their infringement during training.

Defendants' only response is to cite a passage from *Tremblay v. OpenAI, Inc.*, No. 23-cv-03223, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024), which involves entirely different allegations and arguments. *See* OpenAI Supp. Br. at 20.<sup>18</sup> The court held that the plaintiffs cited no "caselaw to suggest that failure to reveal [which internet books it uses to train ChatGPT] has any bearing on whether the alleged removal of CMI in an internal database will knowingly enable infringement." *Tremblay*, 2024 WL 557720, at \*4. But The Intercept relies on factual allegations, not caselaw. It does not allege that Defendants fail to specify which articles it used in general, but whether a given prompt derived from a certain article. And *Tremblay*'s holding is about enablement, while Defendants cite it for concealment. The case is inapposite.

***Concealment from users of infringing outputs.*** The FAC next alleges that removal of CMI during training will conceal the infringement wrought by ChatGPT's outputs. Defendants plausibly know these outputs infringe; ChatGPT itself admits that even partial regurgitation violates copyright. FAC ¶ 66. Due to ChatGPT's serial plagiarism, Defendants had reasonable grounds to know that it would produce infringing outputs. And ChatGPT's failure to include The Intercept's CMI on those outputs—which results from Defendants' removal of CMI during training—plainly conceals that those outputs belong to The Intercept. FAC ¶ 70.

Defendants offer four unpersuasive responses.

---

infringement" by defendants or third parties); *Murphy v. Millennium Radio Grp. LLC*, No. 08-cv-1743, 2015 WL 419884, \*5 (D.N.J. Jan. 30, 2015) (holding that a resulting infringement need not occur).

<sup>18</sup> Microsoft argues that The Intercept "ventures no theory for how the absence of CMI would conceal such infringement, whom it would conceal it from, or how the presence of CMI would reveal infringing conduct during training that would otherwise be kept under wraps." Microsoft Supp. Br. at 8. The above answer all those questions.

First, Defendants dispute that the FAC alleges the requisite knowledge at the time of the removal, pointing out that the FAC cites only recent sources where they admit to that knowledge. OpenAI Supp. Br. at 21. But when Defendants *published* what they knew is different from when they *learned* it. When a party first learned something, or had reasonable grounds to know it, are classic scienter questions requiring discovery. *See, e.g., In re Flag Telecom Holdings, Ltd. Sec. Litig.*, 236 F.R.D. 177, 183 (S.D.N.Y. 2006) (ordering discovery into “when the defendants learned certain facts”). Plus, the FAC also alleges improper removal up to the date of filing, including after the publications Defendants cite. *See* Section II.B.1, *supra*.

Second, Defendants argue that regurgitation of The Intercept’s articles was objectively unlikely *ex ante*. They point to an OpenAI blog post cited in the FAC which says that regurgitations are “more common when particular content appears more than once in training data.” OpenAI and journalism (Jan. 8, 2024), <https://openai.com/index/openai-and-journalism/>. But that comparison means nothing. “Less common” doesn’t mean “unlikely.” 98% likely is less common than 99% likely but is still almost certain.

In any case, contrary to Defendants’ portrayal, *see* OpenAI Supp. Br. at 21-22, the FAC alleges duplication both within and across OpenAI’s training sets, thus making regurgitation *more* likely. As to duplication within the sets, the computer scientists who replicated WebText were forced to de-duplicate the URLs before publishing them online. *See generally* FAC Ex. 2 (referring to URLs as “deduped”). That means there *was* duplication within WebText. As to duplication across training sets, the FAC alleges the presence of Intercept articles in both WebText, and filtered versions of Common Crawl. FAC ¶¶ 41, 43, 57-58. Exhibit 5 to the FAC contains three of The Intercept’s articles as they appear in a filtered version of Common Crawl, and two also appear in the WebText recreation. *Compare* FAC Ex. 5 at 2, 4 *with* FAC Ex. 2 at 7, 20 (listing



identical URLs). Thus, at least on OpenAI’s theory, regurgitation of The Intercept’s articles was *more likely* to have occurred.

Third, Defendants argue that the regurgitations referenced in FAC Exhibit 6 are too *de minimis* to constitute infringement. *See* OpenAI Supp. Br. at 22. But as discussed above in connection with jurisdictional discovery, the FAC plausibly alleges that Defendants’ AI products were far more capable of regurgitation years ago than they are now. *See* Section II.A.2, *supra*.

Fourth, Defendants argue that the FAC fails to plausibly allege that excluding CMI from the training sets would cause ChatGPT to omit CMI from its outputs—even though the FAC alleges exactly that. *See* OpenAI Supp. Br. at 23; *but see* FAC ¶ 70 (“If ChatGPT and Copilot were trained on works of journalism that included the original author, title, copyright notice, and terms of use information, they would have learned to communicate that information when providing responses to users unless Defendants trained them otherwise.”). That well-pleaded allegation is supposedly contradicted by the fact that The Intercept’s regurgitations followed a prompt that, Defendants say, asked ChatGPT to omit CMI. *See* OpenAI Supp. Br. at 23. But the prompt, which merely asked for a “continuation” of the article, did not ask ChatGPT not to omit any particular type of CMI. *See* FAC Ex. 6 at 1. Indeed, copyright notices and terms of use are contained at the end of the article and hence count as continuations. FAC ¶ 46. More fundamentally, though, how The Intercept constructed its prompts has no bearing on how other users interacted with ChatGPT years ago—facts solely within Defendants’ possession. Answering this question, too, requires discovery.

***Inducing, enabling, or facilitating users to infringe.*** The FAC next alleges that Defendants knew, or had reasonable grounds to know, that removing CMI would induce, enable, or facilitate users to infringe The Intercept’s copyright by distributing infringing works to a future

audience. Defendants know users distribute ChatGPT responses this way—they even advertise ChatGPT for that purpose. FAC ¶ 83. Defendants know their users are capable of infringing, including when the users deploy their products as advertised—they have even committed to indemnifying their commercial clients in just these circumstances. FAC ¶ 86. And Defendants know, or at least have reason to know, that removing CMI from training data—causing it not to be included in outputs—would induce at least some users to infringe. That is because at least some people respect copyright or fear copyright liability. FAC ¶¶ 84-85.<sup>19</sup>

Defendants respond with the well-worn objection that a user is unlikely to produce the regurgitations The Intercept attached to its Complaint. *See* Microsoft Supp. Br. at 8-9. The response is the same: ChatGPT is far less prone to regurgitation now than it was years ago.

The Intercept plausibly alleges scienter.

#### **4. The Intercept plausibly alleges the Section 1202(b)(3) claims.**

The Intercept’s Section 1202(b)(3) claims allege that OpenAI and Microsoft distributed the training data, from which CMI had been removed, with each other. FAC ¶¶ 98-99; 110-11. The main dispute concerns whether the FAC alleges that distribution. It does. Microsoft’s CEO, Satya Nadella, said, “we have the data,” giving rise to the plausible inference that Microsoft does, in fact have the data, especially in conjunction with Microsoft’s close partnership with OpenAI, its massive investment in the company, and its provision of the data center and bespoke supercomputing infrastructure that powers ChatGPT. FAC ¶¶ 20-23. And if Microsoft has the

---

<sup>19</sup> This dispenses with Microsoft’s subsidiary arguments about what The Intercept supposed has not explained. Why would an end-user be “inclined to make a further use of an undescribed, unattributed natural-language output”? Microsoft Supp. Br. at 9. Because Defendants advertise it for that purpose. Why would that use be “infringing—versus a fair use”? *Id.* Because Defendants themselves believe their users violate copyright when deploying their AI products as used. Why would the user “not make such a use if it observed CMI”? *Id.* at 9-10. Because they respect copyright, fear liability, or both.

data, which OpenAI played a part in developing, then it plausibly got the data from OpenAI. So OpenAI and Microsoft exchanged the data.

Microsoft reads Mr. Nadella's quote differently as saying that Microsoft has the "rights" to the data, not the data itself. Microsoft Supp. Br. at 6. True, Mr. Nadella mentions rights elsewhere in the interview. But in the same *sentence*, he also says, "we have the people, we have the compute." *Intelligencer Staff, Satya Nadella on Hiring the Most Powerful Man in AI, Intelligencer*, (Nov. 21, 2023), <https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-on-hiring-sam-altman.html>. This suggests Mr. Nadella was talking about what Microsoft physically has, not what it rightfully has. At worst, Mr. Nadella's statement is ambiguous, and ambiguities at the dismissal stage must be resolved in The Intercept's favor. *See Shultz v. Congregation Shearith Israel of City of New York*, 867 F.3d 298, 302 (2d Cir. 2017). Regardless, distribution would be plausible even without this interview. *See P. Opp.* at 23-24.

According to the OpenAI Defendants, meanwhile, the FAC's most "glaring deficienc[y]" is that it does not specify "when, why, or how" the distribution occurred. OpenAI Supp. Br. at 24. But stating a claim requires no such details. *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."). The OpenAI Defendants do not reckon with *Pilla* even though The Intercept cited it in prior briefing. *See P. Opp.* at 2, 24. Their argument is insubstantial.

The OpenAI Defendants next complain that the FAC supposedly does not specify which copies were shared, which it says makes it impossible to assess identity. *See OpenAI Supp. Br.* at 25. But The Intercept has identified the works at issue, FAC Ex. 1, alleged that Defendants shared copies from which CMI had been removed, FAC ¶¶ 75-76, and explained that OpenAI's

CMI removal resulted in copies “substantively identical” to the original but for the absence of CMI, FAC ¶¶ 51-52; FAC Exs. 3-4. Identity is not required, *see ADR Int’l Ltd. v. Inst. for Supply Mgmt. Inc.*, 667 F. Supp. 3d 411, 427-31 (S.D. Tex. 2023), but even if it were, The Intercept has alleged it.<sup>20</sup>

Last, the OpenAI Defendants deny that The Intercept has alleged the second scienter element: that they knew, or had reason to know, that their distribution of CMI-less articles would conceal their infringement. *See* OpenAI Supp. Br. at 25.<sup>21</sup> But clearly the recipient of an unattributed wall of text, *see, e.g.*, FAC Exs. 3-5, would have no idea that it was an unauthorized copy of The Intercept’s works. For all the recipient knows, the original owner—unknown to the recipient—willingly put it in the public domain, or the distributor licensed it. So distributing that unattributed text conceals the infringement.

The Intercept states a Section 1202(b)(3) claim.

**5. The Intercept plausibly alleges Microsoft’s involvement in the training process.**

Finally, The Intercept has plausibly alleged Microsoft’s involvement in the training process, allegations grounding both Counts III and IV of the FAC. Microsoft “ha[s] the data” for a reason, and that reason, plausibly, is to train AI models. FAC ¶ 73. Its relationship with OpenAI, its massive investment in the company, and its provision of bespoke supercomputing resources buttress that conclusion and raise a plausible inference that Microsoft has engaged in AI training similar to that of its close partner. FAC ¶ 74.

---

<sup>20</sup> Defendants do not argue that the trivial differences caused by the text extraction algorithms, such as the random addition of a space between two words, renders the resulting copy non-identical from the original even under their own theory. *See* FAC ¶¶ 51-52.

<sup>21</sup> They do not dispute the first scienter element, and it is simple. If Defendants intentionally removed the CMI (which, except as to Microsoft-specific issues discussed below, they do not dispute the FAC alleges), then they necessarily distributed it “knowing that copyright management information has been removed.” 17 U.S.C. § 1202(b)(3).

Microsoft's responses fail largely for reasons already given. It complains of the lack of detail about the "how or when" of its activities. *But see Pilla*, 2020 WL 1309086, at \*12 (rejecting the need for "how" or "when" at the pleading stage). And it advances a reading of Mr. Nadella's interview that relies on inferences improperly drawn against The Intercept at the pleading stage. *See* Section II.B.4, *supra*. Microsoft also reprises arguments made in its first brief, which fail for reasons discussed elsewhere. *See* P. Opp. at 17.

The Intercept does grant Microsoft one thing: the FAC is fuller on details about OpenAI. The Intercept culled those details largely from public admissions by a younger, formerly nonprofit company with "Open" in its name and different past objectives—much to its present chagrin. But Microsoft has always had the incentive and the business sense to keep similar details secret. So The Intercept relies on OpenAI's public statements, Mr. Nadella's admission to possessing the data, Microsoft's massive investment in OpenAI, its provision of infrastructure, and the companies' close relationship. If, contrary to plausible inferences, Microsoft was not involved, it can prove that in discovery. That should be the next step in this case.

### **III. CONCLUSION**

The Court should deny the motions to dismiss.

/s/ Stephen Stich Match

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: July 15, 2024