

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

THE INTERCEPT MEDIA, INC.,

Plaintiff,

v.

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

Defendants.

Case No. 1:24-cv-01515-JSR

**DEFENDANT MICROSOFT CORPORATION'S SUPPLEMENTAL MEMORANDUM
IN SUPPORT OF ITS MOTION TO DISMISS THE FIRST AMENDED COMPLAINT**

July 8, 2024

TABLE OF CONTENTS

TABLE OF AUTHORITIES iii

PRELIMINARY STATEMENT 1

ARGUMENT 3

I. THE INTERCEPT LACKS STANDING BECAUSE IT FAILS TO ALLEGE AN ACTUAL OR THREATENED INJURY. 3

II. THE INTERCEPT FAILS TO STATE A DMCA § 1202 CLAIM..... 5

 A. The FAC Does Not (And Cannot) Plausibly Allege That *Microsoft* Removed CMI Or Distributed Works Lacking CMI. 5

 B. The FAC Does Not Plausibly Allege A Likelihood That Removal Of CMI During Training Of An AI Tool Will Induce, Enable, Facilitate, Or Conceal Infringement..... 8

CONCLUSION..... 10

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Andersen v. Stability AI Ltd.</i> , No. 23-cv-201-WHO, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023)	10
<i>Ashcroft v. Iqbal</i> , 556 U.S. 662 (2009).....	9
<i>Bell Atlantic v. Twombly</i> , 550 U.S. 544 (2007).....	8
<i>Cartoon Network LP, LLLP v. CSC Holdings, Inc.</i> , 536 F.3d 121, 131 (2d Cir. 2008).....	7
<i>Como v. Commerce Oil Co.</i> , 607 F. Supp. 335 (S.D.N.Y. 1985)	7
<i>Doe 1 v. GitHub, Inc.</i> , No. 22-cv-06823-JST (N.D. Cal. June 24, 2024)	9, 10
<i>Gilliam v. Am. Broad. Cos.</i> , 538 F.2d 14 (2d Cir. 1976).....	5
<i>Lujan v. Defenders of Wildlife</i> , 504 U.S. 555 (1992).....	3
<i>Saba Cap. Cef Opportunities 1, Ltd. v. Nuveen Floating Rate Income Fund</i> , 88 F.4th 103 (2d Cir. 2023)	5
<i>Stevens v. Corelogic, Inc.</i> , 899 F.3d 666 (9th Cir. 2018)	8
<i>TransUnion LLC v. Ramirez</i> , 594 U.S. 413 (2021).....	3, 4
<i>Tremblay v. OpenAI, Inc.</i> , No. 23-cv-03223, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024).....	8, 10
Statutes	
Copyright Act, 17 U.S.C. §§ 101 et seq.	
§ 107.....	9
§ 1202.....	5, 10

§ 1202(b).....1, 2, 8, 10
§ 1202(c).....4

Other Authorities

U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral
Rights in the United States* (April 2019).....5

PRELIMINARY STATEMENT

The Intercept was afforded a chance to “rectify ... the seeming lack of specificity” in its initial Complaint, ECF No. 81 at 1, but the Amended Complaint makes clear that The Intercept does not face a specificity problem—its problem is reality. In particular, two inconvenient realities stand in the way of The Intercept ever alleging a viable § 1202(b) claim. First, the GPT-based tools have never generated, and no facts plausibly suggest they are likely ever to generate, outputs matching The Intercept’s copyrighted works. And second, Microsoft was not involved in creating the datasets underlying the training process from which The Intercept’s § 1202(b) claims arise, and therefore should never have been named as a Defendant in this case. Because the Amended Complaint fails to establish standing or state a claim against Microsoft, the two counts against Microsoft should be dismissed with prejudice.

On the question of whether the accused GPT-tools—OpenAI’s ChatGPT or Microsoft’s “Copilot AI”—have or will output The Intercept’s copyrighted material, The Intercept admitted at oral argument that it sued before even attempting to find out. June 3, 2024 Hrg. Tr. (“Hrg. Tr.”) 22:17-23:9. Now it has tried and failed. With respect to ChatGPT, The Intercept fed it a “snippet” consisting of several paragraphs of one of its articles, terminated in the middle of a sentence, then asked ChatGPT to “complete [the snippet] verbatim.” FAC Ex. 6 at 1. In three examples, ChatGPT finished the sentence with a handful of words—a nubbin appended to an Intercept-supplied snippet.

But no user would ever even ask ChatGPT this kind of thing! Such a user would already need to have The Intercept’s article to prompt ChatGPT in that way, and in any event there is no value to a user in a completed half-sentence. The Intercept does not allege otherwise and fails to offer any example of output of its works under realistic conditions. As for Microsoft’s Copilot tool, The Intercept seemingly nabbed not even a nubbin, providing no examples of Copilot

outputs. That no GPT-based tool has emitted any of The Intercept’s works leaves The Intercept without standing to pursue § 1202(b) claims. Without an output, there is no public dissemination of a CMI-less copy of The Intercept’s works; without public dissemination, there is no concrete injury. The removal of CMI or distribution of CMI-less works during behind-the-scenes training of the GPT models causes no harm. *Infra* § I.

The Intercept also has uncovered no basis for suing *Microsoft* for removal of CMI from its works. Because The Intercept can allege no output of its works, it brings a case solely based on the creation of datasets to train the GPT tools. The initial Complaint attempted to connect Microsoft to training by alleging, “upon information and belief,” that Microsoft “created Common Crawl datasets” in connection with training. Compl. ¶ 31. The Intercept removes this dubious assertion from its Amended Complaint. Meanwhile, it adds various new assertions about how *OpenAI* allegedly created “WebText” datasets, using software called “Dragnet and Newspaper” that extracts only an article’s main text. FAC ¶¶ 45-58. Whatever these allegations are worth in a claim against OpenAI, The Intercept does not and cannot allege them against Microsoft. Its § 1202(b) claims rest on nothing but the fact of Microsoft’s general collaboration with OpenAI and provision of computing and cloud technology, a legally insufficient basis for liability under § 1202(b). *Infra* § II.A.

Finally, the absence of any output of The Intercept’s works—especially when coupled with Microsoft’s lack of involvement with the training process—makes it impossible for The Intercept to plausibly establish § 1202(b)’s scienter requirement. It is objectively implausible that removal of CMI could ever facilitate or conceal copyright infringement because the GPT-based tools are unlikely to ever reproduce The Intercept’s content *at all*. Certainly it is not plausible that *Microsoft*—which was not involved in creating training datasets for the GPT-based

tools and whose own tool is not alleged to have emitted The Intercept's works—had any reason to believe that removal of CMI would facilitate piracy. *Infra* II.B.

The Court should dismiss the claims against Microsoft with prejudice.

ARGUMENT

I. THE INTERCEPT LACKS STANDING BECAUSE IT FAILS TO ALLEGE AN ACTUAL OR THREATENED INJURY.

As explained previously, The Intercept lacks standing because the mere removal of CMI in a non-public setting, without public dissemination of the CMI-stripped copy, causes no concrete injury, and therefore cannot support standing. Mot. 8-13; Reply 2-5. The Amended Complaint fails to solve this defect.

The Amended Complaint does not plausibly allege any real-world public dissemination, via any GPT-based tool, of any copy of an Intercept work without CMI. Quite the opposite, it concedes that when The Intercept tried to get ChatGPT to “regurgitate[] material,” ChatGPT refused. FAC ¶ 66. Even when The Intercept entered multi-paragraph prompts that no user would employ, all The Intercept got was a dozen-or-so words finishing a sentence. FAC, Ex. 6. The notion that this has or would ever happen in the real world, without The Intercept's self-infliction, is beyond “conjectural” and “hypothetical”—it is far-fetched. *See Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992). And even if it did happen, it is hard to see how ChatGPT supplying the last few words to a user who already has the first five paragraphs could be “remotely harmful,” *TransUnion LLC v. Ramirez*, 594 U.S. 413, 426 (2021) (citation omitted).

Previously, The Intercept has pointed to its allegation that Microsoft has removed CMI and then “shared copies of Plaintiffs' works ... with the OpenAI Defendants” to argue that dissemination *between* the Defendants would cause injury. Opp. 23-24; *see* FAC ¶ 75. This argument is audacious because The Intercept fails to offer a single substantive allegation

suggesting that Microsoft was involved in creating datasets for GPT *at all*—let alone involved so comprehensively to support the conclusion that Microsoft must have obtained The Intercept’s works, stripped CMI, then sent them to OpenAI. But even if this speculation were indulged, dissemination between the Defendants is not the sort of *public* lack of attribution that may cause injury. It would be like if the defendant credit agency in *TransUnion* transferred an inaccurate file to another credit reporting agency; if neither “provide[d] those plaintiffs’ credit information to any potential creditors,” there would still be no cognizable harm. 594 U.S. at 430, 433.

Because The Intercept will never be able to allege the sort of public dissemination necessary to claim injury from lack of attribution, it tries to analogize its injury to copyright infringement instead. It claims that the injury it suffers from non-public removal of CMI “is the interference with [its] property.” Hrg. Tr. 13:18-19. If The Intercept is claiming that the CMI itself is the “property” interfered with, that cannot work because CMI is nothing like intellectual property. It is merely attribution “information conveyed in connection with copies ... of a work.” 17 U.S.C. § 1202(c); Reply 3-4.

If instead The Intercept is arguing that its copyrighted articles are the property, and that removal of CMI interferes with *that* property interest, it is simply begging the question of whether CMI removal is indeed analogous to the sorts of invasions of copyright interests cognizable historically or at common law, *TransUnion*, 594 U.S. at 424-25. If a neighbor looked askance at your untended lawn, you could not claim Article III injury merely by touting “interference with property” as a common law analogue. A plaintiff needs to articulate a property-based *injury* that is historically cognizable. Here The Intercept cannot do so because there is no historical basis for treating the mere removal of a title or obscuring of authorship information as interfering with copyright interests. Just the opposite, American law traditionally

rejected *droit morale*—the moral rights theories that would recognize injury based on harm to attribution or expressive integrity. See Mot. 10; Reply 2-4; see U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral Rights in the United States* (April 2019). Simply put, The Intercept has not and cannot identify any historical basis for treating the removal of information about a copyright work as a cognizable *interference with* such a work.

All of this distinguishes this case from the Second Circuit’s decision in *Saba Cap. Cef Opportunities I, Ltd. v. Nuveen Floating Rate Income Fund*, 88 F.4th 103 (2d Cir. 2023), which The Intercept says supports standing here. Hrg. Tr. 12:20-13:6. *Saba* involved corporate action depriving shareholders of voting rights, which were “analogous to property interests.” 888 F.4th at 115. There, unlike here, the plaintiff suffered “concrete harm” because the defendant’s actions “encumbered” its “shares’ voting rights,” thereby “impair[ing] the shares’ ‘condition, quality, or value.’” *Id.* at 116. That injury was “analogous to a property-based injury like conversion or trespass to chattels.” *Id.* But CMI is not intellectual property, and its bare removal has never been understood to invade copyright interests. Only public dissemination of works without CMI could conceivably result in the sort of “presentation of ... work to the public in a distorted form” or “false impression of ... origin” that might qualify as injury, *Gilliam v. Am. Broad. Cos.*, 538 F.2d 14, 24 (2d Cir. 1976). Because The Intercept cannot allege such dissemination, this Court should dismiss the suit for lack of standing.

II. THE INTERCEPT FAILS TO STATE A DMCA § 1202 CLAIM.

A. The FAC Does Not (And Cannot) Plausibly Allege That *Microsoft* Removed CMI Or Distributed Works Lacking CMI.

The Amended Complaint also fails to state a claim against Microsoft because it lacks any plausible allegations that *Microsoft* engaged in the removal of CMI or distribution of works lacking CMI. The Amended Complaint contains no specific allegation substantiating the how or

when of Microsoft allegedly obtaining The Intercept’s works; removing CMI from them; using them to train any GPT-based model or tool; or even possessing copies of The Intercept’s works. Instead, The Intercept bases its claims against Microsoft on three allegations and invites the Court to draw incredibly tenuous suppositions from them.

Quote from Satya Nadella. New to the Amended Complaint is The Intercept’s repeated invocation of Microsoft CEO Satya Nadella’s statement in an interview that Microsoft could “continue the innovation” if “OpenAI disappeared” because “we have the data, we have everything.” FAC ¶ 23. The Intercept then extrapolates from this that Microsoft “has created, without Plaintiff’s permission, its own copies of Plaintiff’s copyright-protected works of journalism.” FAC ¶ 73. Even if The Intercept were fairly quoting Mr. Nadella, the statement “we have the data” hardly supports the conclusion that Microsoft made copies of The Intercept’s works, let alone the further necessary inference that Microsoft *removed CMI* from The Intercept’s works.

But The Intercept is not fairly quoting him. The partial quote omits the full context from the article, which the Amended Complaint incorporates by reference.* What Mr. Nadella says is that Microsoft has “all of the rights” to continue OpenAI’s work. He then goes on to explain precisely what Microsoft does in connection with development of the GPT-based products: “We do the kernel optimizations, we build tools, we build the infrastructure.” *Id.* In context, the quote says only that Microsoft has provided technology and infrastructure to OpenAI and that its investment in OpenAI confers certain “rights” to continue the innovation in which Microsoft has invested.

* *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>.

Provision of Technology. The Intercept next tries to draw inferences from allegations that Microsoft “created and hosted the data centers used to develop ChatGPT,” FAC ¶ 74, and “provi[ded] database and computing resources,” FAC ¶ 75. From this provision of technology, The Intercept leaps directly to conclusory allegations that “Microsoft intentionally removed” CMI and “shared copies of Plaintiff’s works.” FAC ¶¶ 74-76. These conclusions do not follow. As Microsoft has explained, it is well-settled in the copyright context that providing someone the technology and infrastructure that is used to commit some act is not the same thing as committing that act oneself (and certainly not *intentionally*). Mot. 19-20 (citing *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 131 (2d Cir. 2008)). The Intercept must plausibly allege facts showing that *Microsoft itself* removed CMI, not that someone else did so using a Microsoft computer. And tellingly, The Intercept *does* include allegations that purport to explain how OpenAI has actually removed CMI from works for purposes of creating GPT training sets, but *does not* claim that Microsoft engaged in that conduct. It invokes the provision of technology because it has nothing else to go on.

Relationship with OpenAI. What remains of the Amended Complaint is The Intercept’s guilt-by-association theory—that whatever it has alleged against OpenAI must also be true of Microsoft because of “the relationship” between the companies. As Microsoft has explained (at Mot. 19), “[t]he fact that two companies ... do work together” is insufficient “to attribute their acts to one another.” *Como v. Commerce Oil Co.*, 607 F. Supp. 335, 340 (S.D.N.Y. 1985). And that is all the Amended Complaint alleges: that because Microsoft and OpenAI have a collaboration, Microsoft must have been involved in the highly technical process of LLM training and thus must have removed CMI from Intercept works during such training or distributed such works. That is far from enough to “allow[] a potentially massive factual

controversy to proceed,” *Bell Atlantic v. Twombly*, 550 U.S. 544, 558 (2007) (quotation marks omitted), which is why district courts are empowered to require plaintiffs to “rectify ... [a] lack of specificity” in a complaint, ECF No. 81 at 1, before greenlighting a lawsuit. Microsoft should not have to defend against liability on the basis of The Intercept’s non-existent allegations about Microsoft’s involvement in LLM training.

B. The FAC Does Not Plausibly Allege A Likelihood That Removal Of CMI During Training Of An AI Tool Will Induce, Enable, Facilitate, Or Conceal Infringement.

Finally, the Amended Complaint fails to state a § 1202(b) claim because The Intercept does not allege a likelihood of infringement of its works, as necessary to establish Microsoft’s awareness that removal of CMI from a work in a training set will somehow “induce, enable, facilitate, or conceal” such infringement under 17 U.S.C. § 1202(b). Mot. 20-23; Reply 8-10.

The Intercept does not dispute that a § 1202(b) claim requires it to establish that it is “likely from an ex ante perspective” that CMI removal will somehow conceal or facilitate infringement. Hrg. Tr. 22:4-5, 21-22; *see Stevens v. Corelogic, Inc.*, 899 F.3d 666, 675 (9th Cir. 2018). It seems to advance two theories, both entirely conclusory. One is that removal of CMI will “conceal copyright infringement by Defendants,” FAC ¶ 109, ostensibly referring to alleged infringement during training, *see* FAC ¶¶ 77-81 (allegations concerning OpenAI’s conduct during training). But 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. *See Tremblay v. OpenAI, Inc.*, No. 23-cv-03223, 2024 WL 557720, at *3-4 (N.D. Cal. Feb. 12, 2024).

The Intercept’s other theory is output-based, so it is a non-starter for reasons already explained. The Intercept says that “inclusion in training sets of Plaintiff’s works of journalism without [CMI] would induce ChatGPT and Bing AI products to provide responses to users that

incorporated material from Plaintiff’s copyright-protected works or regurgitated copyright-protected works verbatim or nearly verbatim.” FAC ¶ 105. Then, the theory continues, users will “distribute or publish responses” they receive, which they “would not have” done if CMI were never removed during training. FAC ¶ 106.

As already explained, the best The Intercept could get ChatGPT to “regurgitate” was a handful of words that completed a four- or five-paragraph excerpt The Intercept itself supplied. So its theory is built on a string of suppositions that become more and more farcical as you read them. First, someone randomly comes upon five full continuous paragraphs of an Intercept article with the last sentence cut off. Then, not knowing that it is an Intercept article, this person feeds it to ChatGPT and asks for the remainder of the sentence. ChatGPT returns some words to complete the sentence, but does not provide The Intercept’s CMI. This imaginary user then further distributes the second half of the sentence *as a result of* ChatGPT not providing the CMI.

Plausibility is about “*reasonable inference[s].*” *Ashcroft v. Iqbal*, 556 U.S. 662, 663 (2009) (emphasis added). The Amended Complaint is a house of cards. Nothing in the Amended Complaint explains why any real-world user with access to an article would prompt a GPT-based tool in the way The Intercept has. Its allegations are nothing but lawyer contrivances necessitated by the basic reality that the AI tools at issue simply do not output The Intercept’s works. *See Order, Doe 1 v. GitHub, Inc.*, No. 22-cv-06823-JST (N.D. Cal. June 24, 2024), ECF No. 253 at 5-6 (CMI claims must be based on GPT-based tools in “*normal operation*”) (emphasis added). And even if there was some likelihood that one of the GPT-based tools would output The Intercept’s works, The Intercept still fails to allege why (a) an end-user would be inclined to make a further use of an undescribed, unattributed natural-language output; (b) that use would be infringing—versus a fair use, 17 U.S.C. § 107; and (c) why a user so inclined would not make

such a use if it observed CMI. Without such an account, there is no reason to believe that removing CMI to train an AI model would facilitate piracy, as necessary for a § 1202(b) claim.

Courts have repeatedly and decisively rejected § 1202 challenges to LLM-based tools for a simple reason: Those tools typically do not output identical copies of plaintiffs' works at all, making it impossible to plead removal of CMI from such copies or likely piracy caused by such removal. *Doe 1*, No. 22-cv-06823-JST (N.D. Cal. June 24, 2024), ECF No. 253, at 4-6; *Tremblay*, 2024 WL 557720, at *4-5; *Andersen v. Stability AI Ltd.*, No. 23-cv-201-WHO, 2023 WL 7132064, at *10-11 (N.D. Cal. Oct. 30, 2023). In the face of this ever-growing line of decisions, The Intercept sought a loophole. It would avoid even finding out if a GPT-based tool ever would output its works without CMI, focusing only on conduct during their *training*.

But the workaround does not work. It does not work because the non-public removal of CMI does not injure The Intercept. It does not work because without outputs matching The Intercept's works, there is no reason to believe removal of CMI will in any way facilitate or conceal piracy—the ill to which § 1202(b) is directed. And when it comes to Microsoft, the theory does not work because The Intercept cannot allege that Microsoft actually engaged in the removal of CMI or distribution of CMI-stripped works that allegedly took place at the training stage.

CONCLUSION

The Court should grant the motion to dismiss in its entirety and dismiss the claims against Microsoft with prejudice.

Dated: July 8, 2024

Orrick, Herrington & Sutcliffe LLP

San Francisco, CA

By: /s/ Annette L. Hurst

Annette L. Hurst (admitted *pro hac vice*)
The Orrick Building
405 Howard Street
San Francisco, CA 94105-2669
Telephone: (415) 773-5700
ahurst@orrick.com

Lisa T. Simpson
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3767
lsimpson@orrick.com

Christopher J. Cariello
51 West 52nd Street
New York, NY 10019
Telephone: (212) 506-3778
ccariello@orrick.com

*Attorneys for Defendant
Microsoft Corporation*