

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

THE CENTER FOR INVESTIGATIVE
REPORTING, 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.

Case No. 1:24-cv-04872-SHS

**DEFENDANT MICROSOFT CORPORATION'S MEMORANDUM
IN SUPPORT OF MOTION TO DISMISS THE COMPLAINT**

September 3, 2024

TABLE OF CONTENTS

TABLE OF AUTHORITIES ii

PRELIMINARY STATEMENT 1

ALLEGATIONS OF THE COMPLAINT..... 3

A. Microsoft Collaborates With OpenAI To Bring GPT-Based Products To The Public..... 3

B. CIR Sues Microsoft Based On Alleged Participation In Creating Or Distributing Datasets Used To Train GPT-Based Products..... 4

C. The Complaint Alleges Copyright Infringement and CMI Removal Claims Against Microsoft, but Points To Neither Concrete Harm Nor Any Instance Of GPT-Based Tools Outputting Its Works..... 5

ARGUMENT 8

I. CIR LACKS STANDING FOR ITS CMI CLAIMS BECAUSE IT FAILS TO ALLEGE AN ACTUAL OR THREATENED INJURY. 8

A. Absent Dissemination, Removal Of CMI In A Non-Public Setting Inflicts No Concrete Harm Under Article III. 9

B. CIR Lacks Standing To Seek Damages And Injunctive Relief For CMI Claims..... 12

II. CIR FAILS TO STATE A § 1202 CLAIM. 13

A. The Complaint Does Not Plausibly Allege That Microsoft Removed CMI From Copies Of Works Or Distributed Works Lacking CMI. 14

B. The Complaint Does Not Plausibly Allege A Likelihood That Removal Of CMI During Training Of An AI Tool Will Induce, Enable, Facilitate, Or Conceal Infringement..... 17

III. CIR FAILS TO STATE A CLAIM FOR CONTRIBUTORY INFRINGEMENT BASED ON OUTPUTS GENERATED BASED ON USER PROMPTS. 20

A. The Complaint Fails To Allege An Actionable Instance Of End-User Copyright Infringement..... 21

B. The Complaint Fails To Allege Microsoft’s Knowledge Of Direct Infringement. 23

CONCLUSION..... 25

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Alan Ross Mach. Corp. v. Machinio Corp.</i> , No. 17-cv-3569, 2019 WL 1317664 (N.D. Ill. Mar. 22, 2019)	13
<i>Andersen v. Stability AI Ltd.</i> , 700 F. Supp. 3d 853 (N.D. Cal. 2023).....	20
<i>Andersen v. Stability AI Ltd.</i> , No. 23-cv-00201, 2024 WL 3823234 (N.D. Cal. Aug. 12, 2024).....	1
<i>Ashcroft v. Iqbal</i> , 556 U.S. 662 (2009)	8, 17, 18
<i>Bell Atlantic Corp. v. Twombly</i> , 550 U.S. 544 (2007)	17
<i>BMG Rights Mgmt. (US) LLC v. Cox Commc’ns, Inc.</i> , 881 F.3d 293 (4th Cir. 2018)	23, 24
<i>Burress v. Freedom Mortg. Corp.</i> , No. 20-cv-15242, 2022 WL 2916070 (D.N.J. July 22, 2022)	10
<i>Carter v. Helmsley-Spear, Inc.</i> , 71 F.3d 77 (2d Cir. 1995).....	11
<i>Cartoon Network LP, LLLP v. CSC Holdings, Inc.</i> , 536 F.3d 121 (2d Cir. 2008)	16
<i>Clapper v. Amnesty Int’l USA</i> , 568 U.S. 398 (2013)	13
<i>Como v. Commerce Oil Co.</i> , 607 F. Supp. 335 (S.D.N.Y. 1985)	17
<i>Daily News, LP v. Microsoft Corp.</i> , No. 24-cv-03285 (S.D.N.Y. Apr. 30, 2024)	<i>passim</i>
<i>Dinerstein v. Google, LLC</i> , 73 F.4th 502 (7th Cir. 2023).....	10
<i>Doe I v. GitHub, Inc.</i> , No. 22-cv-06823, 2024 WL 235217 (N.D. Cal. Jan. 22, 2024)	1
<i>Fashion Nova, LLC v. Blush Mark, Inc.</i> , No. 22-cv-6127, 2023 WL 4307646 (C.D. Cal. June 30, 2023).....	12

Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,
499 U.S. 340 (1991) 21

Free Speech Sys., LLC v. Menzel,
390 F. Supp. 3d 1162 (N.D. Cal. 2019)..... 20

Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc.,
443 F.2d 1159 (2d Cir. 1971)..... 21, 23

Gilliam v. Am. Broad. Cos.,
538 F.2d 14 (2d Cir. 1976) 11

Kent v. Universal Studios, Inc.,
No. 08-cv-2703, 2008 WL 11338293 (C.D. Cal. Aug. 15, 2008)..... 11

Luvdarts, LLC v. AT&T Mobility, LLC,
710 F.3d 1068 (9th Cir. 2013)..... 24, 25

Mango v. BuzzFeed, Inc.,
970 F.3d 167 (2d Cir. 2020) 17, 18

Matthew Bender & Co. v. West Publ'g Co.,
158 F.3d 693 (2d Cir. 1998)..... 21, 22

MGM Studios Inc. v. Grokster, Ltd.,
545 U.S. 913 (2005) 25

Ringgold v. Black Ent. Television, Inc.,
126 F.3d 70 (2d Cir. 1997) 22

Roberts v. BroadwayHD LLC,
518 F. Supp. 3d 719 (S.D.N.Y. 2021) 12

Sony Music Ent. v. Cox Commc'ns, Inc.,
93 F.4th 222 (4th Cir. 2024)..... 24

Spokeo, Inc. v. Robins,
578 U.S. 330 (2016) 9

Steel v. Bongiovi,
784 F. Supp. 2d 94 (D. Mass. 2011)..... 13

Stevens v. Corelogic, Inc.,
899 F.3d 666 (9th Cir. 2018)..... 18, 20

Suid v. Newsweek Mag.,
503 F. Supp. 146 (D.D.C. 1980)..... 11

Tilford v. Jones,
No. 05-cv-2989, 2006 WL 2612752 (S.D. Tex. Sept. 11, 2006)..... 11

Town of Chester, N.Y. v. Laroe Ests., Inc.,
581 U.S. 433 (2017) 9

TransUnion LLC v. Ramirez,
594 U.S. 413 (2021) *passim*

Tremblay v. OpenAI, Inc.,
Nos. 23-cv-03223, 23-cv-03416, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024)..... *passim*

VanderKodde v. Elliott,
No. 17-cv-203, 2023 WL 9898588 (W.D. Mich. Apr. 27, 2023)..... 10

Viacom Int’l, Inc. v. Youtube, Inc.,
676 F.3d 19 (2d Cir. 2012) 24

Statutes

Copyright Act, 17 U.S.C. §§ 101 et seq.

§ 106..... 21

§ 106A 11

§ 107 6

§ 501 5

§ 512(c)..... 24

§ 1202(b) *passim*

§ 1202(b)(1)..... 6, 13

§ 1202(b)(3)..... 7, 13, 14

§ 1203(a)..... 13

Federal Rules of Civil Procedure

Rule 12(b)(1) 2, 8, 9, 13

Rule 12(b)(6) 8, 13

Other Authorities

Intelligencer Staff, *Satya Nadella on Hiring the Most Powerful Man in AI* (Nov. 21, 2023), <https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-on-hiring-sam-altman.html> 15

S. Rep. 105-190 (1998)..... 12

U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral Rights in the United States* (April 2019)..... 12

PRELIMINARY STATEMENT

The Center for Investigative Reporting (CIR) brings the latest tag-along lawsuit challenging Microsoft and OpenAI’s groundbreaking generative AI tools. Like plaintiffs before it, CIR alleges that its works were part of the “training sets” for OpenAI’s GPT models, and it asserts that using its works to develop LLM technology constitutes copyright infringement. That claim will ultimately need to confront copyright’s fair use doctrine, among other defenses. But also like many of its predecessors, CIR overreaches, asserting that when the public uses the transformational AI tools Defendants offer, this somehow also harms CIR’s rights or even threatens the very existence of its industry. None of these claims has ever been substantiated by factual allegations about how actual real-world people use generative AI tools. And CIR’s Complaint offers nothing new, except the latest confirmation that these claims of harm are empty.

This motion seeks to dismiss three claims. The first two are claims under § 1202(b) of the Digital Millennium Copyright Act, alleging that Microsoft removed copyright management information (CMI) from CIR’s works or distributed works with the CMI removed. Over the past year, courts have repeatedly rejected § 1202 claims against LLM-based tools for a simple reason: because those tools generate new content on the fly, no plaintiff has yet been able to allege, much less prove, that an identical copy of an entire work from the training set was ever output, making it impossible to plead removal of CMI from such copies.¹ So CIR tries a different way: It alleges not that CMI is removed from *outputs* of GPT-based tools, but that Defendants removed

¹ *Tremblay v. OpenAI, Inc.*, Nos. 23-cv-03223, 23-cv-03416, 2024 WL 557720, at *5 (N.D. Cal. Feb. 12, 2024); *Doe I v. GitHub, Inc.*, No. 22-cv-06823, 2024 WL 235217, at *9 (N.D. Cal. Jan. 22, 2024); *id.*, No. 22-cv-06823 (N.D. Cal. June 24, 2024), ECF No. 253 at 4-6 (dismissal with prejudice); *Andersen v. Stability AI Ltd.*, No. 23-cv-00201, 2024 WL 3823234, at *8 (N.D. Cal. Aug. 12, 2024).

CMI during non-public *training*—and that this removal alone supports a claim.

The gambit fails several times over. Most fundamentally, CIR’s attempt to predicate a § 1202 claim on mere non-public removal of CMI leaves it unable to allege the sort of concrete injury necessary for standing under Article III. To have standing for a claim, a plaintiff must point to a concrete and legally cognizable injury. The mere removal of CMI from a work, without public dissemination of the work, results in no such injury—at most, CIR alleges there is a bare technical violation of a statute causing no real harm. Because CIR alleges nothing more, its CMI claims should be dismissed under Rule 12(b)(1). *Infra* § I.

Even if CIR did have standing, it nevertheless fails to state a § 1202(b) claim against Microsoft. While it purports to predicate a claim on removal of CMI during the training of OpenAI’s GPT models, it offers no plausible allegations that *Microsoft* engaged in the downloading of works, collection of such works into a training set, or training itself. Without that, there is no basis for asserting that Microsoft removed CMI from CIR’s works. CIR’s claim also fails for the independent reason that the Complaint fails to allege that removal of CMI from its works during training is objectively likely to conceal or facilitate infringement—a requirement for any § 1202(b) claim. Indeed, CIR cannot point to a single real-world instance of any user *ever* using a prompt that would elicit an output that infringes CIR’s works, let alone explain why this user would be likely to further infringe CIR’s works because of absent CMI. Each of these independent reasons merits dismissal. *Infra* § II.

CIR’s claim for contributory copyright infringement also fails. CIR’s theory here is that if end-users infringe via their prompting of a GPT-based tool, Defendants should be held liable for that infringement. But stating such a claim requires CIR to allege both that a specific instance of infringement has actually occurred *and* that Defendants had actual knowledge of

those specific acts. Again, the Complaint contains not a single instance of any real-world user ever using a GPT-based tool to infringe CIR’s works. At best, CIR alleges that the GPT-based products are capable, in a highly-stylized setting, of emitting trivial snippets of CIR works or bullet-point summaries of articles with direct citation to the original. Even if this were enough to state a claim for direct infringement, it comes nowhere close to showing that Microsoft *knew* of these instances of infringement and materially contributed to them. Quite the opposite, the Complaint acknowledges that both Microsoft and OpenAI take steps to *prevent* any possible infringement—which is presumably why CIR can point to none that has actually happened. The contributory infringement claim should also be dismissed. *Infra* § III.

ALLEGATIONS OF THE COMPLAINT

A. Microsoft Collaborates With OpenAI To Bring GPT-Based Products To The Public.

Over the past several years, scientists and engineers have made dramatic advances in the field of artificial intelligence. The “large language model,” or LLM, is one of those advances. LLMs extract the elements of language by “ingest[ing]” information based on large bodies of data called “training sets.” Compl. ¶¶ 5, 47. Once trained, an LLM can generate natural-language responses to user prompts, “mimic[king] how humans write and speak.” Compl. ¶ 5.

According to the Complaint, “Defendants are companies responsible for the creation and development of ... ChatGPT and Copilot artificial intelligence (AI) products,” Compl. ¶ 4—products built on GPT models that “provide responses to questions or other prompts” from users, Compl. ¶ 76. The Complaint asserts that Microsoft “invested billions of dollars” into OpenAI and “provide[d] the data center and ... supercomputing infrastructure used to train ChatGPT” and “created and hosted the data centers used to develop ChatGPT.” Compl. ¶¶ 27-28, 102. The Complaint also alleges “Microsoft has built its own AI product, called Copilot, which uses

Microsoft’s Prometheus technology.” Compl. ¶ 51. In turn, “Prometheus combines the Bing search product with the OpenAI Defendants’ GPT models into a component called Bing Orchestrator,” a product that likewise provides natural-language responses to user inquiries. Compl. ¶ 51.

B. CIR Sues Microsoft Based On Alleged Participation In Creating Or Distributing Datasets Used To Train GPT-Based Products.

The Center for Investigative Reporting is a “diversified multi-media nonprofit organization” that “runs, *inter alia*, the brands Mother Jones, Reveal and CIR Studios.” Compl. ¶¶ 2, 12. It alleges that Defendants included CIR works in the training set when developing GPT-based products, and in doing so “intentionally removed author, title, copyright notice, and terms of use information from [CIR’s] copyrighted works.” Compl. ¶ 102; *see* Compl. ¶¶ 103-04.

The Complaint acknowledges that CIR does not have direct knowledge of what is contained in the training sets for the GPT models because “Defendants have not published the contents of the training sets used to train any version of ChatGPT.” Compl. ¶ 50. It bases the allegation that “thousands of [CIR’s] copyrighted works were included in Defendants’ training sets” on “publicly available information” and the view of its “data scientist.” Compl. ¶ 99. It asserts that “versions of ChatGPT ... were trained using at least the following training sets: WebText, WebText2, and sets derived from Common Crawl.” Compl. ¶ 52; *see also* Compl. ¶¶ 70-72, 99. The Complaint contains no allegations about the datasets used to train Microsoft’s “Bing Copilot” or “Bing AI.”

The Complaint also repeatedly asserts that CMI was removed from CIR’s works, but offers no direct allegations of how or when this happened during the training process and who was involved in that conduct. It instead relies on alleged technological generalizations. The

Complaint asserts that “[i]f ChatGPT and Copilot were trained on works of journalism that included the [CMI], they would have learned to communicate that information when providing responses to users unless Defendants trained them otherwise.” Compl. ¶ 98.

Most starkly, the Complaint also lacks any non-conclusory factual allegations that *Microsoft* specifically removed CMI from CIR’s works in creating datasets, or that *Microsoft* specifically distributed CIR’s works knowing that the CMI had been removed from those works. As to the former, the Complaint alleges “WebText and WebText2 were created by the OpenAI Defendants,” and, obliquely, that “*Defendants* have a record, and are aware, of each URL that was included in each of their training sets.” Compl. ¶¶ 53, 55 (emphasis added). Beyond that, the Complaint points to an incomplete quote from Microsoft’s CEO that “we have the data, we have everything,” along with the allegation that “Microsoft created and hosted the data centers used to develop ChatGPT.” Compl. ¶¶ 101-02. From this the Complaint claims that “Microsoft intentionally removed [CMI] from Plaintiff’s copyrighted works in creating ChatGPT and Copilot training sets,” and that “Microsoft has shared copies” of CIR’s works without CMI. Compl. ¶¶ 101-03.

C. The Complaint Alleges Copyright Infringement and CMI Removal Claims Against Microsoft, but Points To Neither Concrete Harm Nor Any Instance Of GPT-Based Tools Outputting Its Works.

The Complaint alleges four causes of action against Microsoft. The first is Count II, for direct copyright infringement under 17 U.S.C. § 501, which alleges that Microsoft “download[ed]” CIR’s copyrighted works from the internet, Compl. ¶ 126; “encod[ed]” them “in computer memory,” Compl. ¶ 127; “regurgitat[ed] those works verbatim or nearly verbatim,” Compl. ¶ 128; and “ma[d]e ... cop[ies] ... in the memory of their computing system” to create an “abridgment” of works, Compl. ¶ 85. Because of a low standard needed to properly plead this

claim, it is not at issue in this motion, but will ultimately turn, among other things, on a fair-use defense under 17 U.S.C. § 107 not suited to the pleading stage.

At issue in this motion are Counts III, VI, and VII. Count III alleges contributory infringement against Microsoft and Open AI, anticipating a defense to the direct infringement claims—that it is the user, not Microsoft or OpenAI, who is the direct infringer if it is the “user[’s] prompts” that generate the allegedly infringing output. Compl. ¶ 134. The Complaint alleges that “to the extent a user may be liable as a direct infringer based on output of ChatGPT and/or Copilot, Defendants materially contributed to and directly assisted with the direct infringement by those users.” Compl. ¶ 133. The Complaint provides no instance of alleged infringement of its works by a user. Yet it claims that “Defendants knew or had reason to know” of infringement because “Defendants undertake extensive efforts in developing, testing, or troubleshooting their models” and “have agreed to defend and indemnify certain of their users for copyright violations only when the users are using the products according to terms specified by Defendants.” Compl. ¶ 134.

The Complaint’s sixth cause of action alleges violation of § 1202(b)(1) of the DMCA, which protects against the removal of “copyright management information,” or CMI, from copies of works. The Complaint alleges that Microsoft “created copies of Plaintiff’s works ... with” “author, title, copyright notice, and terms of use information” “removed and included them in training sets used to train ChatGPT and Bing AI products.” Compl. ¶¶ 149-53. It further alleges that Microsoft “had reason to know” that inclusion in training sets of CIR’s works without CMI would “induce” users to “distribute or publish” copyrighted works “that such users would not have distributed or published if they were aware of” the CMI, and that this would “enable,” “facilitate,” and “conceal” copyright infringement. Compl. ¶¶ 154-57. The

Complaint’s seventh cause of action alleges violation of § 1202(b)(3) of the DMCA, alleging that Microsoft “shared copies of Plaintiff’s works without author, title, copyright notice, and terms of use information with the OpenAI Defendants in connection with the development of ChatGPT and Copilot.” Compl. ¶ 159.

As far as the Complaint discloses, CIR has suffered no concrete harm as a result of the alleged removal of CMI from its “thousands” of works. Compl. ¶ 99. The closest it comes is a conclusory allegation that attempts to satisfy § 1202(b)’s scienter requirement, which demands that the defendant intended, knew, or should have known that removal of CMI would facilitate or conceal infringement. The Complaint posits that if “ChatGPT or Copilot provid[es] responses to ChatGPT users that abridge[] or regurgitate[] material verbatim from copyrighted works,” those users might “further distribute the results,” but “would be less likely to [do so] if they were made aware of the author, title, copyright, and terms of use information.” Compl. ¶¶ 111-13.

Nor does the Complaint allege any real-world examples in which ChatGPT or Copilot has ever emitted anything similar to or derived from its works. It cites “[e]xamples of three ... regurgitations” that CIR elicited by an elaborate prompt, resulting in outputs of two sentences or less. Compl. ¶ 81. The Complaint also alleges that CIR itself prompted summaries of Mother Jones articles and received such summaries as output. Compl. ¶ 89 (citing Ex. 8). Aside from this, the Complaint notes that the plaintiffs in a *different* case, *Daily News v. Microsoft Corp.*, claimed that they were able to use their own creative prompting to elicit “regurgitations” of the *Daily News* plaintiffs’ works. Compl. ¶ 78 (citing Compl. Ex. J, *Daily News, LP v. Microsoft Corp.*, No. 24-cv-03285 (S.D.N.Y. Apr. 30, 2024)). But the Complaint simultaneously claims that OpenAI has “recently changed ChatGPT to reduce regurgitations for copyright reasons,”

Compl. ¶ 80, while elliptically acknowledging that Microsoft has deployed “guardrails and content filters” to prevent the end-user infringement CIR fails to allege, Compl. ¶ 115.

ARGUMENT

To survive a motion to dismiss under Rule 12(b)(1) based on a lack of Article III standing, “a plaintiff must show (i) that [it] suffered an injury in fact that is concrete, particularized, and actual or imminent; (ii) that the injury was likely caused by the defendant; and (iii) that the injury would likely be redressed by judicial relief.” *TransUnion LLC v. Ramirez*, 594 U.S. 413, 423 (2021). Whether an injury is concrete for purposes of Article III depends on whether the “plaintiff[] ha[s] identified a close historical or common-law analogue for their asserted injury”; that inquiry cannot depend “on contemporary, evolving beliefs about what kinds of suits should be heard in federal courts.” *Id.* at 424-25.

To survive a motion to dismiss under Rule 12(b)(6), “a complaint must contain sufficient factual matter, accepted as true, to state a claim to relief that is plausible on its face.” *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (quotation marks and citation omitted). If a claim lacks a cognizable theory, lacks sufficient facts to plausibly support that theory, or advances a theory that is foreclosed as a matter of law, dismissal is appropriate. “Threadbare recitals of the elements of a cause of action, supported by mere conclusory statements,” cannot state a claim. *Id.*

I. CIR LACKS STANDING FOR ITS CMI CLAIMS BECAUSE IT FAILS TO ALLEGE AN ACTUAL OR THREATENED INJURY.

CIR lacks standing for its claims under § 1202(b) of the DMCA because the Complaint fails to plausibly allege any actual or threatened injury. *See* Compl. ¶¶ 148-59 (Counts VI and VII). CIR’s DMCA claims are based solely on alleged removal of CMI during the process of training the GPT models—that is, they are *not* predicated on alleged real-world public dissemination of works lacking CMI. The mere removal of CMI in a non-public setting from a

copy that never sees the light of day causes no harm. The Court should dismiss CIR’s CMI-removal claims under Rule 12(b)(1) for lack of standing.

A. Absent Dissemination, Removal Of CMI In A Non-Public Setting Inflicts No Concrete Harm Under Article III.

“Standing is not dispensed in gross. To the contrary, a plaintiff must demonstrate standing for each claim [it] seeks to press and for each form of relief that is sought.” *Town of Chester, N.Y. v. Laroe Ests., Inc.*, 581 U.S. 433, 439 (2017) (cleaned up). To establish standing, a plaintiff must demonstrate an injury for each claim that is both “concrete and particularized and actual or imminent, not conjectural or hypothetical.” *Spokeo, Inc. v. Robins*, 578 U.S. 330, 339 (2016) (citation omitted). That means a concrete injury “must actually exist”; it cannot be “speculative.” *Id.* at 340. And it must be a “real [injury], and not [an] abstract [one],” *id.*: While “Congress may ‘elevate’ harms that ‘exist’ in the real world before Congress recognized them to actionable legal status, it may not simply enact an injury into existence, using its lawmaking power to transform something that is not remotely harmful into something that is.” *TransUnion*, 594 U.S. at 426 (citation omitted). Only those harms with “a close historical or common-law analogue for the[] asserted injury” can confer standing. *Id.* at 424.

The Supreme Court applied these principles in *TransUnion v. Ramirez*. The plaintiffs in *TransUnion* “alleged violations of the Fair Credit Reporting Act” for failing “to follow reasonable procedures to ensure the accuracy of credit files so that the files would not include [Office of Foreign Assets Control] alerts labeling the plaintiffs as potential terrorists.” 594 U.S. at 430. For a subset of these plaintiffs, defendant TransUnion did indeed maintain records with “misleading OFAC alerts,” but had never “provide[d] those plaintiffs’ credit information to any potential creditors.” *Id.* at 433. All “assumed” that TransUnion had committed a statutory

violation. *Id.* at 432 n.5. The question was whether these plaintiffs had suffered a concrete injury merely based on the non-public maintenance of false information about them.

The Supreme Court held that they had not. The Court explained that “[t]he mere presence of an inaccuracy in an internal credit file, if it is not disclosed to a third party, causes no concrete harm.” *Id.* at 434. Reasoning that “[p]ublication” is essential to a claim for defamation—the closest common-law analogue to a claim based on false credit information—the Court held that “no historical or common-law analog” supports an injury based on “the mere existence of inaccurate information.” *Id.* Nor could those plaintiffs gain standing by restyling their injury as a “material risk of future harm” that inaccurate information might be disseminated. *Id.* at 435. Harm must “materialize” to support a “suit for damages.” *Id.* at 436. And even to maintain a suit for injunctive relief, a plaintiff must show that the risk of harm is “imminent and substantial.” *Id.* at 435. Courts have applied *TransUnion*’s framework to find lack of concrete injury in various contexts.²

The CMI claims here are remarkably similar to those that failed in *TransUnion*. They invoke a statute of recent vintage that for the first time endows a type of information—CMI—with legal import, then creates a novel legal violation for its removal from a copy. CIR at most claims nothing but a bare technical violation of that statute. It points to no instance in which copies of its works lacking CMI were ever disseminated, nor substantiates the idea that there is imminent and substantial risk of non-attribution of its copyrighted works to the public. To the contrary, CIR admits that when it tried to get ChatGPT to “regurgitat[e]” copies, it refused.

² *Dinerstein v. Google, LLC*, 73 F.4th 502, 512-16 (7th Cir. 2023) (finding lack of Article III injury for state-law privacy claim based on use of anonymized patient records to create AI healthcare product); *VanderKodde v. Elliott*, No. 17-cv-203, 2023 WL 9898588, at *5 (W.D. Mich. Apr. 27, 2023) (no Article III injury underlying alleged violations of the Fair Debt Collection Practices Act); *Burress v. Freedom Mortg. Corp.*, No. 20-cv-15242, 2022 WL 2916070, at *4 (D.N.J. July 22, 2022) (same, for alleged violations of the Truth in Lending Act).

Compl. ¶ 80. Even when CIR entered multi-paragraph prompts that no user would employ, the most CIR got was a sentence or two to complete the prompt. Compl., Ex. 7. And even if a real-world user did ask for a couple more sentences of an article the user already possessed, it is hard to see how ChatGPT supplying that information would be “remotely harmful.” *TransUnion*, 594 U.S. at 426 (citation omitted). The only alleged harm here is thus the private *non-attribution* of CIR’s *ownership* of certain works—which is to say, no harm at all.

Nothing in the common law remotely suggests otherwise. In general, the mere lack of attribution—even public non-attribution—“is not a cognizable cause of action under ... common law.” *Tilford v. Jones*, No. 05-cv-2989, 2006 WL 2612752, at *4 (S.D. Tex. Sept. 11, 2006); *see also Suid v. Newsweek Mag.*, 503 F. Supp. 146, 149 (D.D.C. 1980) (noting that the court had “been unable to locate[] any case recognizing a common-law action for failure to attribute”). American courts have at times found rights of attribution in “the tort of unfair competition” or “state unfair competition laws.” *Gilliam v. Am. Broad. Cos.*, 538 F.2d 14, 24 (2d Cir. 1976); *but see Kent v. Universal Studios, Inc.*, No. 08-cv-2703, 2008 WL 11338293, at *4 (C.D. Cal. Aug. 15, 2008) (finding *Gilliam* overruled by *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23 (2003)). But even if such a theory were viable, it would be based on “the author’s personal right to prevent the presentation of his work *to the public* in a distorted form” or to prevent “a false impression of ... origin.” *Gilliam*, 538 F.2d at 24 (emphasis added); *see also Carter v. Helmsley-Spear, Inc.*, 71 F.3d 77, 81 (2d Cir. 1995) (recognizing that the “civil law” right of attribution is “*personal to the artist*”) (emphasis added); *cf.* 17 U.S.C. § 106A (narrow right of action for human “authors” of visual works to “claim authorship”).

Not only is there no common-law tradition suggesting that private non-attribution of ownership information constitutes an injury—Congress also would not have viewed it as an

actionable statutory injury when it enacted § 1202(b). The aim of the DMCA was to “discourage piracy,” S. Rep. 105-190, at *11 n.18 (1998); “assist in tracking and monitoring uses of *copyrighted works*,” and foster an “efficient Internet marketplace,” *id.* at *16 (emphasis added). “[T]he purpose of CMI is to provide *the public* with notice that a work is copyrighted.” *Fashion Nova, LLC v. Blush Mark, Inc.*, No. 22-cv-6127, 2023 WL 4307646, at *5 (C.D. Cal. June 30, 2023); see *Roberts v. BroadwayHD LLC*, 518 F. Supp. 3d 719, 737 (S.D.N.Y. 2021) (same). The reason for protecting CMI is not some freestanding interest in having attribution information attached to a copy of a work, but to prevent harms from *public dissemination* of that work without CMI. Removal of CMI without dissemination of a copy of a work therefore does not result in even the sort of injury Congress had in mind when it enacted the DMCA.

Nor can CMI removal be analogized to an interference with CIR’s copyright-protected property, i.e., its articles. 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 U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral Rights in the United States* (April 2019).

B. CIR Lacks Standing To Seek Damages And Injunctive Relief For CMI Claims.

The Complaint’s failure to allege injury leaves CIR without standing to pursue either damages or injunctive relief for CMI claims. With respect to damages, CIR would need to plausibly allege that some injury has “materialize[d],” *TransUnion*, 594 U.S. at 436—that is, that a GPT-based product has actually disseminated a copy of its works without CMI or that a GPT-based product has contributed to an end-user’s infringement. As explained above, it has done neither.

As for injunctive relief, CIR seeks “[a]n injunction requiring Defendants to remove all copies of [CIR’s] copyrighted works from which author, title, copyright, or terms of use information was removed from their training sets and any other repositories.” Compl. at 32, ¶ (iii). Insofar as the only injury for which CIR seeks injunctive relief is the mere presence of works in training sets, that is not a redressable injury for the reasons already explained. It may be that CIR also means to predicate standing for injunctive relief on the purported risk that Microsoft may at some point disseminate a copy of one of CIR’s works without CMI. If so, the Complaint proves that such risk is nothing but conjecture based on lawyer-contrived prompts that no real-world user would ever use. At any rate, courts should be “reluctan[t] to endorse standing theories that rest on speculation about the decisions of independent actors.” *Clapper v. Amnesty Int’l USA*, 568 U.S. 398, 414 (2013). CIR’s nascent theory is far too speculative to confer standing based on a risk-of-future-harm injury. *Id.* at 401 (A future injury is speculative if it is not “certainly impending.”).

II. CIR FAILS TO STATE A § 1202 CLAIM.

In addition to failing for lack of standing under Rule 12(b)(1), CIR’s CMI-removal claims should also be dismissed under Rule 12(b)(6). Compl. ¶¶ 148-59 (Counts VI and VII). At the threshold, for reasons just explained, CIR is not a “person injured” and therefore cannot bring a DMCA claim. 17 U.S.C. § 1203(a); *see Steel v. Bongiovi*, 784 F. Supp. 2d 94, 97-98 (D. Mass. 2011); *Alan Ross Mach. Corp. v. Machinio Corp.*, No. 17-cv-3569, 2019 WL 1317664, at *4 (N.D. Ill. Mar. 22, 2019). CIR also fails to state a claim against Microsoft under either § 1202(b)(1) or (b)(3).

To state a claim under § 1202(b)(1), a plaintiff must allege that the defendant “intentionally remov[ed] or alter[ed]” CMI from a copy of the plaintiff’s work. 17 U.S.C. § 1202(b)(1). To state a claim under § 1202(b)(3), a plaintiff must plausibly allege that the

defendant “distribute[d]” copies of the plaintiff’s works “knowing that [CMI] has been removed or altered without authority of the copyright owner[.]” 17 U.S.C. § 1202(b)(3). And for both claims, a plaintiff must further establish that the defendant removed CMI or distributed CMI-less copies “knowing, or ... having reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement of any right under this title.” 17 U.S.C. § 1202(b). Here, CIR’s attempt to plead these claims against Microsoft fails because the Complaint does not plausibly allege that *Microsoft* ever removed CMI from copies of CIR’s works or distributed copies lacking CMI. *Infra* § A. CIR’s § 1202 claims also fail because the Complaint does not plausibly allege that any purported removal of CMI or distribution of works would have an objective likelihood of facilitating infringement or thwarting efforts to police it. *Infra* § B.

A. The Complaint Does Not Plausibly Allege That Microsoft Removed CMI From Copies Of Works Or Distributed Works Lacking CMI.

The Complaint 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 during training of the GPT models. Rather, it alleges merely that *OpenAI*—in creating OpenAI’s GPT models—engaged in various activities to assemble datasets and train OpenAI’s models, then incants the words “upon information and belief” to claim that Microsoft must have done these same things. That is insufficient to plausibly allege that Microsoft removed CMI from CIR’s works.

The contrast between the allegations against OpenAI and Microsoft is stark. The Complaint offers twenty paragraphs of allegations about how *OpenAI* allegedly scraped data, created “WebText” datasets, used software called “Dagnet and Newspaper” that extracts only an article’s main text, and ultimately trained its models. Compl. ¶¶ 53-73. Whatever these allegations are worth in a claim against OpenAI, CIR does not and cannot allege a single one of

them against Microsoft. The most the Complaint alleges about Microsoft's own actions is that, "upon information and belief," Microsoft "shared copies of Plaintiff's works without author, title, copyright notice, and terms of use information *with the OpenAI Defendants* in connection with the development of ChatGPT and Copilot." Compl. ¶ 159 (emphasis added); *see* Compl. ¶ 103. But the Complaint contains no factual allegations describing what Microsoft supposedly shared with OpenAI, when or how it purportedly did so, or why doing so would even have been necessary.

Without direct or specific allegations, CIR bases its § 1202 claims against Microsoft on mere association with OpenAI. It rests this theory on three empty allegations.

First, CIR invokes 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." Compl. ¶¶ 29, 101. CIR claims that this quote demonstrates that "Microsoft has created, without Plaintiff's permission, its own copies of Plaintiff's copyright-protected works of journalism, including but not limited to the Registered Works." Compl. ¶ 101. Even if CIR were fairly quoting Mr. Nadella, the statement "we have the data" hardly supports the conclusion that Microsoft made copies of CIR's works, let alone the further necessary inference that Microsoft *removed CMI* from those works.

In any event, the partial quote CIR provides omits the full context from the article from which it is taken, which the Complaint incorporates by reference (Compl. ¶ 29 & n.2).³ What Mr. Nadella says is that Microsoft has "all of the rights" to continue OpenAI's work. He goes on to explain precisely what Microsoft does in connection with development of the GPT-based

³ 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>.

products: “We do the kernel optimizations, we build tools, we build the infrastructure.” *Supra* n.3. With this 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. As described below, *infra* 17-20, that is not enough to state a § 1202 claim.

Second, CIR alleges that Microsoft “created and hosted the data centers used to develop ChatGPT,” Compl. ¶ 102, and provided “database and computing resources to the OpenAI Defendants,” Compl. ¶ 103. From this provision of technology, CIR leaps directly to conclusory allegations that “Microsoft intentionally removed” CMI and “shared copies of Plaintiff’s works.” Compl. ¶¶ 102-03. These conclusions do not follow.

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*). *Cartoon Network LP, LLLP v. CSC Holdings, Inc.*, 536 F.3d 121, 131 (2d Cir. 2008); 3 Patry on Copyright § 9:5.50 (2024) (explaining that volitional conduct requires that a “defendant had to have chosen to engage in the conduct that is adjudged to be infringing” and noting that it is “lacking where a defendant’s system is merely used to create a copy by a third party”) (citation omitted). Where a defendant does nothing more than “design[], house[], and maintain[] a system” that someone else uses to make a copy or initiate a distribution, the defendant has not engaged in that conduct. *Cartoon Network*, 536 F.3d at 131. This principle forecloses the Complaint’s contention that Microsoft, by hosting works, actively and knowingly distributes those works. Were it otherwise, every cloud storage provider or DVR manufacturer would be automatically liable for distributing works without CMI any time its equipment automatically copied or transmitted a copy made by someone else. Nothing in

§ 1202(b)'s text, structure, or purpose supports such sweeping liability. Without a plausible allegation that Microsoft itself removed CMI—not that someone else did so using a Microsoft computer—CIR's claims against Microsoft fail.

Third and finally, CIR falls back on Microsoft's "working relationship" with OpenAI. Compl. ¶ 104. But "[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). CIR's suggestion that because Microsoft and OpenAI have a collaboration, Microsoft must have been involved in the highly technical process of LLM training is far from enough "allow[] a potentially massive factual controversy to proceed." *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544, 558 (2007) (citation omitted). Microsoft should not have to defend against liability on the basis of CIR's "[t]hreadbare" allegations about Microsoft's involvement in OpenAI's LLM training. *Iqbal*, 556 U.S. at 678.

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

CIR's § 1202 claims fail for the additional reason that the Complaint does not plausibly allege that removal of CMI during training would be likely to "induce, enable, facilitate, or conceal an infringement." 17 U.S.C. § 1202(b). Because it is an anti-piracy statute, § 1202 is careful not to impose liability on a defendant whose removal of CMI has only the incidental effect of aiding or concealing copyright infringement. It enacts this limitation through a stringent "double-scienter requirement." *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 172 (2d Cir. 2020) (capitalization altered). The first half of the requirement demands that the defendant's conduct—removal of CMI, alteration of CMI, and so forth—be intentional (for subsection (b)(1)) or knowing (for subsection (b)(3)). *See id.* The second half, at issue here, requires that the

“defendant know or have reason to know” that its conduct “will induce, enable, facilitate, or conceal an infringement.” *Id.*

Though this latter requirement is phrased as a scienter standard, courts have recognized that it carries an objective component in cases involving technology that may incidentally remove CMI. Unless a plaintiff can point to direct facts revealing the defendant’s subjective belief that removal of CMI will facilitate infringement, it must plausibly allege that, as a result of technology’s removal of CMI, infringement is *objectively* “likely.” *Stevens v. Corelogic, Inc.*, 899 F.3d 666, 675 (9th Cir. 2018). Otherwise, there is no basis for inferring that a defendant “was aware or had reasonable grounds to be aware” that “the probable future impact of its actions” would be infringement. *Id.* at 674.

CIR fails to plausibly allege this objective likelihood of future infringement. The Complaint’s first theory is that removal of CMI at the training stage “would conceal copyright infringement by Defendants.” Compl. ¶ 157. The Complaint 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*, 2024 WL 557720, at *3-4.

The Complaint’s second theory asserts in conclusory fashion that Defendants “had reason to know that inclusion in training sets of [CIR’s] works” without CMI “would enable copyright infringement by ChatGPT, Bing AI, and ChatGPT and Bing AI users” (Compl. ¶ 155), as well as “facilitate” (Compl. ¶ 156) and “conceal” (Compl. ¶ 157) infringement by those users. But beyond these “[t]hreadbare recitals of” the “element[] of [the] cause of action,” *Iqbal*, 556 U.S. at 678, the Complaint offers nothing but unsubstantiated speculation about how the GPT-based tools work and how users of them behave, *see* Compl. ¶¶ 105-15.

According to CIR, the GPT-based products are likely to generate outputs that “abridge[] or regurgitate[]” CIR’s works, and do not include CMI. Compl. ¶ 110. But the Complaint offers no instance of this ever happening in the real world. Indeed, it does not even offer a plausible *user prompt* that would generate such outputs. *See supra* 7, 11; Compl. Ex. 7. Instead, it points only to a handful of examples that it coaxed a GPT-based tool to elicit. The so-called “regurgitations” required CIR to input extensive excerpts of CIR’s own articles, then ask the tool to complete the excerpt—a prompt that yielded two sentences or fewer at the end of an article the user already had. As for the “abridgements,” these are bullet-point summaries of articles that the prompt *requested by name* and that come with a *direct link* to the article. It is dubious that these constitute copyright infringement at all, let alone that they support likely infringement by users. *Infra* 21-23.

The Complaint then claims that Defendants “had reason to know that users of ChatGPT would further distribute the results of ChatGPT responses” (thus, supposedly, furthering infringement) because “ChatGPT [is] a tool that can be used by a user to generate content for a further audience.” Compl. ¶ 112. But again, that is utterly implausible based on the allegations in this Complaint. When CIR used its own contrived prompts to try to get a GPT-based tool to emit output matching CIR’s works, the best it could coax was a couple-dozen words completing a sentence or a bullet-point summary. *Supra* 11. The Complaint offers no reason a user would ever want to redistribute something so trivial.

Last, CIR alleges that users “would be less likely to distribute ChatGPT responses if they were made aware” of CMI, on the theory that “at least some” users “respect the copyrights of others or fear liability for copyright infringement.” Compl. ¶ 113. This one makes the least sense of all. Again, CIR’s only purported examples of offending outputs require the user to

already have either the work or the title of the article. The Complaint offers no reason to think that a user that already has all of this information would somehow be moved to behave differently by being supplied additional CMI.

The Ninth Circuit’s decision in *Stevens v. Corelogic* rejected a similarly speculative attempt to claim likely future infringement. *See* 899 F.3d at 675. *Stevens* involved the defendant’s removal of CMI metadata from photographs as part of technological “[d]ownsampling”—*i.e.*, compression. *Id.* at 671. Plaintiff photographers suggested that this somehow would aid infringement. But they “ha[d] not ... averred that they ha[d] ever used CMI metadata to prevent or detect copyright infringement.” *Id.* at 675. Nor could they establish that the defendant was aware of a third-party “pattern of conduct” or “established modus operandi” giving rise to an inference that the defendant was “aware of the probable future impact of its actions” with respect to CMI. *Id.* at 674. The plaintiffs therefore could not establish that any removal of CMI was done with an awareness that it would facilitate infringement. *Id.* at 675. They had “simply identif[ied] a general possibility that exists whenever CMI is removed,” not a likelihood or an awareness that it will occur. *Id.* at 673-75. Courts have not hesitated to apply *Stevens* at the pleading stage, and doing so is appropriate here.⁴ Without any plausible allegations that Microsoft’s actions furthered infringement, the § 1202 claims should be dismissed.

III. CIR FAILS TO STATE A CLAIM FOR CONTRIBUTORY INFRINGEMENT BASED ON OUTPUTS GENERATED BASED ON USER PROMPTS.

CIR also fails to state a claim that Microsoft has contributed to end-user copyright infringement. *See* Compl. ¶¶ 132-34 (Count III). A contributory infringer is “one who, with

⁴ *See Tremblay*, 2024 WL 557720, at *3-4; *Andersen v. Stability AI Ltd.*, 700 F. Supp. 3d 853, 871 (N.D. Cal. 2023); *Free Speech Sys., LLC v. Menzel*, 390 F. Supp. 3d 1162, 1174-75 (N.D. Cal. 2019).

knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another.” *Gershwin Publ’g Corp. v. Columbia Artists Mgmt., Inc.*, 443 F.2d 1159, 1162 (2d Cir. 1971). The Complaint does not allege that Microsoft encouraged end-users to make infringing use of any GPT-based product, acted in a way that would promote such infringing use, or did anything to induce such infringing use. It advances only a “material[] contribut[ion]” theory. Compl. ¶ 133. The notion is that by offering end-users products that it allegedly knows are “capable of distributing unlicensed copies” or “abridgements of” works, Microsoft is a contributory infringer any time a user makes use of that capability. *Id.*

This theory fails for two reasons. First, CIR fails to allege any instance of direct infringement by a user, a necessary prerequisite to a claim for secondary liability. *Infra* § A. Second, CIR cannot allege that Microsoft knowingly participated in any specific infringement. *Infra* § B.

A. The Complaint Fails To Allege An Actionable Instance Of End-User Copyright Infringement.

“Contributory infringement necessarily must follow a finding of direct or primary infringement.” *Matthew Bender & Co. v. West Publ’g Co.*, 158 F.3d 693, 706 (2d Cir. 1998) (quoting *Cable/Home Commc’n Corp. v. Network Prods., Inc.*, 902 F.2d 829, 845 (11th Cir. 1990)). To allege direct infringement, a plaintiff must plead (1) its “ownership of a valid copyright,” and (2) a primary infringer’s “copying of constituent elements of the work that are original”—*i.e.*, infringement of one of the rights protected under 17 U.S.C. § 106. *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 361 (1991).

In *Matthew Bender*, West Publishing Company, the publisher of the “National Reporter System,” sued the publisher of electronic databases that collected judicial opinions. 158 F.3d at 696-97. West alleged that a user of these databases could operate them to “view (and print)

judicial opinions in the same order in which they are printed in a West volume,” thus recreating West’s reporters. *Id.* at 697. On this basis, it sought to hold the “database manufacturer ... liable as a contributory infringer ... for creating a product that assists a user to infringe a copyright directly.” *Id.* at 706. The Second Circuit rejected the claim. Though West “hypothesized that users ... will retrieve and print cases in the order in which they appear in West’s case reporters,” it “failed to identify any primary infringer, other than [its own] counsel.” *Id.*

CIR has the same problem here. The Complaint “hypothesize[s]” that someone *could* find a way to prompt the GPT-based products to yield output that infringes CIR’s works. But nowhere does the Complaint actually allege that any real-life user has done so. Instead, like other plaintiffs before it, CIR tries to itself coax the GPT-based products to yield outputs upon which it can predicate a lawsuit. The Complaint points to “three ... regurgitations,” Compl. ¶ 81, elicited in response to CIR’s own prompt (without any indication of the number of inquiries or prompt variations needed to elicit the three responses). That prompt supplied what CIR calls a “snippet” of an article, but which actually reflects a several-dozen-word excerpt of a work. Compl. Ex. 7 at 2-4. Not only that—apparently for CIR’s trick-prompt to work, the “snippet” must also be cut off in the middle of a sentence. *See id.* Then the prompt instructs the GPT-based product to “complete [the snippet] *verbatim*,” and “(respond with continuation only).” *Id.* at 1.

But the Complaint nowhere alleges that any real person has used or would use a prompt like that. And even when CIR used it, the most it elicited was non-verbatim output of two sentences. *See* Compl. Ex. 7 at 2-4. These tiny bits of text are almost certainly insufficient to even qualify as copyright infringement. *See Ringgold v. Black Ent. Television, Inc.*, 126 F.3d 70, 74-76 (2d Cir. 1997) (recognizing that “trivial” or “de minimis” copying is not actionable). But

this Court need not resolve that question, because there is no plausible allegation that such outputs have ever even happened. The same goes for CIR’s examples of bullet-point summaries of articles, accompanied by a direct link to the “full article.” Compl. Ex. 8. The notion that briefly summarizing an article while providing a direct citation to it constitutes copyright infringement would utterly decimate the public domain. But here there is no plausible allegation of any real-world instance of such a summary anyway, foreclosing any claim.

CIR also cannot fill the gap by pointing to an exhibit filed in *Daily News* based on those plaintiffs’ contrived prompts. Compl. ¶ 78 (citing Compl. Ex. J, *Daily News, LP v. Microsoft Corporation*, No. 24-cv-03285 (S.D.N.Y. Apr. 30, 2024)). To begin with, that exhibit has nothing to do with CIR or its works. But CIR also admits that it “attempted to obtain the same regurgitations set forth in the *Daily News* case using the same methodology,” but was unsuccessful. Compl. ¶ 80. That CIR, deliberately attempting to replicate pre-provided prompts, failed to get a GPT-based tool to provide offending output makes it still more implausible that a user would generate such output prompting in the dark—undermining not just CIR’s claims but the *Daily News* plaintiffs’ as well. See Microsoft’s Mot. to Dismiss, *Daily News, LP v. Microsoft Corp.*, No. 24-cv-03285 (S.D.N.Y. June 11, 2024), ECF No. 77 at 6-9. Because CIR alleges nothing but the bare theoretical possibility of end-user infringement, its claim should be dismissed. See *Tremblay*, 2024 WL 557720, at *2-3 (dismissing vicarious infringement claim for failure to adequately allege direct infringement).

B. The Complaint Fails To Allege Microsoft’s Knowledge Of Direct Infringement.

The Complaint also fails to allege “knowledge of the infringing activity.” *Gershwin*, 443 F.2d at 1162. To state a material-contribution claim, a plaintiff must allege more than mere “[g]eneral[] know[ledge] of infringement,” especially where the defendant offers a product or service with noninfringing uses. *BMG Rights Mgmt. (US) LLC v. Cox Commc’ns, Inc.*, 881 F.3d

293, 311 (4th Cir. 2018). “Selling a product with both lawful and unlawful uses suggests an intent to cause infringement only if the seller knows of *specific* instances of infringement.” *Id.*; accord *Luvdarts, LLC v. AT&T Mobility, LLC*, 710 F.3d 1068, 1072-73 (9th Cir. 2013) (demanding knowledge of “specific acts of infringement”); cf. *Viacom Int’l, Inc. v. Youtube, Inc.*, 676 F.3d 19, 32 (2d Cir. 2012) (knowledge standard under safe harbor provision of DMCA, 17 U.S.C. § 512(c), requires “awareness of facts or circumstances that indicate specific and identifiable instances of infringement”).

The bedrock requirement of knowing participation also demands that the defendant have “actual knowledge” of (or be willfully blind to) specific acts of infringement. *Luvdarts*, 710 F.3d at 1072-73. It is not enough that a defendant “should have known of ... infringing activity.” *BMG*, 881 F.3d at 310. And as a logical matter, the defendant must have knowledge of the specific act before it takes place—that is, “predictive” knowledge—because one cannot knowingly participate in an act learned of only after it is completed. *Sony Music Ent. v. Cox Commc’ns, Inc.*, 93 F.4th 222, 234 (4th Cir. 2024).

The Complaint’s allegations of Microsoft’s knowledge are pure generality. It alleges that Defendants knew or had reason to know of “the direct infringement by their users because ... Defendants undertake extensive efforts in developing, testing, or troubleshooting their models.” Compl. ¶ 134. It also alleges that Defendants “have committed to paying the user’s costs in defending against [an] infringement claim, and to indemnifying the user,” and claims this shows that “Defendants know or have reason to know that ChatGPT and Copilot users *are capable of infringing* and likely to infringe copyright even when used according to terms specified by Defendants.” Compl. ¶ 115 (emphasis added); *see also* Compl. ¶ 134 (similar). But none of this suggests that Microsoft knew or even had some mechanism for knowing of a “specific act[] of

infringement,” *Luvdarts*, 710 F.3d at 1072-73, let alone that it had actual knowledge of such and act before it took place such that one could infer Microsoft’s knowing participation.

Copyright law bars liability based on “the equivocal conduct of selling an item with substantial lawful as well as unlawful uses,” “limit[ing] liability to instances of more acute fault than the mere understanding that some of one’s products will be misused.” *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 932-33 (2005). That rule applies here and requires dismissal.

CONCLUSION

The Court should grant the motion to dismiss in its entirety.

Dated: San Francisco, California
September 3, 2024

Orrick, Herrington & Sutcliffe LLP

By: /s/ Annette L. Hurst

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

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

Sheryl Koval Garko (*Pro Hac Vice* pending)
222 Berkley Street
Boston, MA 02116
Telephone: (617) 880-1800
sgarko@orrick.com

Laura Brooks Najemy (*Pro Hac Vice*)
222 Berkley Street
Boston, MA 02116
Telephone: (617) 880-1800
lnajemy@orrick.com

*Attorneys for Defendant
Microsoft Corporation*