UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

DAILY NEWS, LP; CHICAGO TRIBUNE
COMPANY LLC; ORLANDO SENTINEL
COMMUNICATIONS COMPANY, LLC; SUN-
SENTINEL COMPANY, LLC; SAN JOSE
MERCURY-NEWS, LLC; DP MEDIA
NETWORK, LLC; ORB PUBLISHING, LLC;
AND NORTHWEST PUBLICATIONS, LLC,

                Plaintiffs,

     v.

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

                Defendants.

Case No. 1:24-cv-03285 (SHS) (OTW)

**MEMORANDUM OF LAW
IN SUPPORT OF OPENAI
DEFENDANTS' MOTION TO
DISMISS**

ORAL ARGUMENT REQUESTED

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

CASES

STATUTES

OTHER AUTHORITIES

I.   INTRODUCTION

This is a copycat lawsuit, modeled after the one the New York Times filed against OpenAI in December 2023. Plaintiffs are eight regional and local newspapers (the "Publishers"), largely owned by a private equity fund that reportedly has made billions by gutting local newsrooms.[1] They contend that OpenAI violated their intellectual property rights by using publicly available, years-old newspaper articles as part of the process of developing a series of artificial intelligence tools. The core claim is that this conduct amounts to direct copyright infringement. But the complaint also includes ancillary causes of action as well, for contributory copyright infringement, trademark infringement, common law misappropriation, and a violation of a portion of the Copyright Act that concerns misuse of information about copyrighted works, rather than the works themselves.

This motion principally seeks dismissal of the ancillary claims. The contributory infringement claim, for example, would ascribe liability to OpenAI based on generalized knowledge of third-party infringement, rather than actual knowledge of specific infringements, which the law requires. The claim for violations of 17 U.S.C. § 1202 (the "DMCA"), which is the provision concerning "copyright management information," fails for numerous reasons, including the reasons embraced by every other court to consider indistinguishable claims against generative AI models: the DMCA simply does not address the conduct to which the Publishers seek to ascribe liability. The claim for state common law misappropriation is preempted by the federal Copyright

---

[1] *See* P.M. Abernathy, The State of Local News: The 2022 Report, Northwestern University (June 29, 2022), https://localnewsinitiative.northwestern.edu/research/state-of-local-news/2022/report/ ("State of Local News") (explaining that Alden Capital submitted "unsolicited" but successful bid for the Chicago Tribune in 2021); Lukas I. Alpert & Cara Lombardo, *Tribune Says Alden Wins Approval Amid Confusion Over Key Shareholder's Vote*, The Wall Street Journal (May 21, 2021, 2:58 PM), https://www.wsj.com/articles/tribune-publishing-says-sale-to-alden-global-capital-wins-approval-11621620732 (discussing same sale and explaining that Alden Capital also owns MediaNews Group, which publishes the Denver Post and the San Jose Mercury News).

Act. The trademark claim fails because Plaintiffs' marks are not sufficiently famous to support a claim for dilution. The motion also seeks dismissal of the direct copyright infringement claim to the extent that it asserts liability for conduct outside the time period prescribed in the operative statute of limitations.

This motion does not seek to resolve the core issue of whether using copyrighted content to train a generative AI model is fair use under copyright law. That key question will require the development of record evidence to show, *inter alia*, that Plaintiffs have flagrantly mischaracterized what OpenAI's products are and how they work. To take just one example of many, the suggestion that ChatGPT can be used to generate at will the verbatim text of newspaper articles is obviously and demonstrably false. Plaintiffs' contrary suggestions depend on an elaborate effort to coax such outputs from OpenAI's products, in a way that violates the operative OpenAI terms of service and that no normal user would ever even attempt, combined with a willfully misleading presentation of the results. At the end of the day, the truth will emerge, and it will be clear that ChatGPT is not in fact some highly inefficient way to access, via one out of every thousand or so impermissible attempts, snippets of old newspaper articles that are freely available in full elsewhere online.

But that is for a later stage of the litigation. For now, OpenAI respectfully asks the Court to grant this more narrowly tailored motion, in order to rid the case of legal theories that are infirm on their face, and allow the parties to focus on the core issue.

II.   **BACKGROUND**

        A.   <u>OpenAI And Its GPT Models</u>

                1.   *Language Model Research and the Innovation of Scale*

OpenAI was founded in 2015 to "advance digital intelligence in the way that is most likely to benefit humanity as a whole." Compl. ¶ 53. OpenAI researches and develops "language

models"[2] that can make "predict[ions]" based on statistics derived from a body of text. Compl. ¶ 75.

OpenAI has pioneered major innovations in language models by building models based on training datasets of enormous scale. Researchers created such a dataset in 2019 by compiling text from webpages whose URLs had been publicly shared on a social media platform.[3] They built the dataset this way in part to overcome past issues with documents "whose contents [were] mostly unintelligible." GPT-2 Paper at 3. Using documents that humans had already read allowed researchers to ensure that the data would be of higher "quality." *Id*. The results of this compilation became a dataset called "WebText," which contained "slightly over 8 million documents." *See id*. Because of how the dataset was created—*i.e.*, by drawing from social media platform links posted by users who were primarily English-speaking—some domains whose links were more frequently shared were more frequently represented in the dataset. *See* GPT-2 Paper at 3.

OpenAI used WebText to train a language model called "GPT-2." Compl. ¶ 82. The results were extremely promising: GPT-2 could perform higher-function tasks than previous models, which had been limited by the narrowness of their training data. GPT-2 Paper at 1, 6–7. This showed that scale was the key to language processing. Researchers expanded the WebText database into "WebText2," which included links shared over "a longer period of time."[4] *See* Compl. ¶ 84. Like the original WebText, the dataset reflected links users had shared. Researchers also used a filtered version of Common Crawl, a "copy of the internet," (Compl. ¶ 85), a much

---

[2] Sébastien Bubeck, et al., *Sparks of Artificial General Intelligence: Early Experiments with GPT-4* at 4, 98 (Apr. 13, 2023), https://arxiv.org/pdf/2303.12712.pdf ("Bubeck Paper")); *see also* Compl. ¶ 68 (citing and quoting this paper).
[3] OpenAI, *Language Models are Unsupervised Multitask Learners* at 3 (Feb. 14, 2019), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf ("GPT-2 Paper"); *see also* Compl. ¶ 82 n.22 (citing and quoting this paper).

[4] OpenAI, *Language Models are Few-Shot Learners* at 4 (July 22, 2020), https://arxiv.org/pdf/2005.14165.pdf ("GPT-3 Paper"); *see also* Compl. ¶¶ 83, 87 & nn. 24, 28 (citing and quoting this paper).

larger dataset.[5]   At the time, it was known that WebText2 and Common Crawl included content from Plaintiffs' websites.[6]   The GPT-3 model trained on this dataset could perform even more complex tasks, and was much more flexible. *See* GPT-3 Paper at 5, 22.

        2.   *OpenAI Today*

OpenAI built on its early success by researching, developing, and releasing subsequent models—GPT-3.5 (in 2022) and GPT-4 (in 2023)—which triggered the AI revolution that we are living through today. GPT-4 "can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more." Bubeck Paper at 1.

        a.   *ChatGPT*

Users may interact with OpenAI's models using ChatGPT, a consumer-friendly platform—accessible for free at chat.openai.com—to "chat" with an OpenAI model through a user interface.[7] Users can input prompts into the platform to generate outputs. ChatGPT's "Browse with Bing" feature enables ChatGPT to fetch recent information about events that occurred after the models' training "cutoff," using the Bing search engine. Compl. ¶ 69. After its release in November 2022, ChatGPT became an "instant viral sensation." Compl. ¶ 58. According to the Complaint, the service gained over 100 million users globally in its first three months. *Id.* Those users rely on

---

[5] The filtered Common Crawl dataset cited in OpenAI's paper included 410 billion tokens. *See* GPT-3 Paper at 9. The C4 dataset, "a filtered English-language subset of a 2019 snapshot of Common Crawl" that Plaintiffs cite in the Complaint, has 156 billion tokens—less than half that size. Dodge et al., *Documenting Large Webtext Corpora: A Case Study On the Colossal Clean Crawled Corpus* at 1 (Sept. 30, 2021), https://arxiv.org/pdf/2104.08758. Any representations about the Chicago Tribune or other papers' weight in the C4 set thus have little to say about the representation of the Tribune in OpenAI's filtered Common Crawl set.

[6] *See* Compl. ¶ 85 (noting Tribune content in "2019 snapshot of Common Crawl"); OpenAI, GPT-2 Model Card, Github, https://github.com/openai/gpt-2/blob/master/model_card.md (last updated Nov. 2019) ("GPT-2 Model Card") (listing Plaintiffs' websites); *see also* Compl. ¶ 82 n.21 (citing and quoting this source).

[7] Compl. ¶¶ 59, 96; OpenAI Blog, *Introducing ChatGPT Plus* (Feb. 1, 2023), https://openai.com/blog/chatgpt-plus; *see also* Compl. ¶ 179 n.75 (citing this page).

ChatGPT to "craft clearer communications, accelerate coding tasks, rapidly explore answers to complex business questions, assist with creative work, and much more."[8]

b. *Custom GPTs*

Users can also create "custom GPTs"—tools built by users, rather than by OpenAI itself, that "create a tailored version of ChatGPT" for a "specific purpose."[9] Custom GPTs can use "third party APIs," or developer tools not controlled by OpenAI, in order to "integrate external data or interact with the real-world"—in other words, to reach outside of ChatGPT and bring in data from other sites or sources.[10] The "custom GPT" store is like an app store; it allows users to browse custom GPTs built by others, which may include third-party tools that connect to external sources.[11]

3. *The New York Times' Suit, And The Publishers' Follow-On Suit*

On December 27, 2023, the New York Times sued OpenAI and Microsoft, alleging that both the training and outputs of OpenAI's models infringed its copyrights, unfairly competed with its business, and diluted its brand. *See* Compl. (Dkt. 1), *The New York Times v. Microsoft*, Dkt. 1:23-cv-11195-SHS (filed Dec. 27, 2023). The Times included in its complaint screenshots of several alleged problematic outputs. *See id.*

On April 30, 2024, Plaintiffs filed this suit. *See* Dkt. 1. Just like the New York Times, Plaintiffs take issue both with OpenAI's training of its models (as described above), and the outputs of those models, alleging that both infringe their copyrights, dilute their brands, and unfairly

---

[8] Compl. ¶ 179; OpenAI Blog, *Introducing ChatGPT Enterprise* (Aug. 28, 2023), https://openai.com/blog/introducing-chatgpt-enterprise; *see also* Compl. ¶¶ 60, 179 & nn.8 & 76 (citing this article).
[9] OpenAI Blog, *Introducing GPTs* (Nov 6., 2023) ("Introducing GPTs Blog Post"), https://openai.com/blog/introducing-gpts; *see* Compl. ¶ 147 n. 57 (citing this article).
[10] *Id.*
[11] *Id.*

compete for news consumers. And just like the Times, Plaintiffs include examples of alleged problematic outputs. *See* Dkt. 1-10 (Ex. J).

B.   Plaintiffs Focus on Two Fringe Behaviors: Regurgitation & Hallucination

As to the outputs, Plaintiffs focus on two uncommon and unintended behaviors that they say harm them in distinct ways: regurgitation and hallucination. Regurgitation (which Plaintiffs also refer to as "memorization," Compl. ¶ 77) occurs when a language model "generat[es] a sample that closely resembles [its] training data."[12] While rare, this most often happens "[w]hen the training data set contains a number of highly similar observations, such as duplicates" of a particular work. Burg Paper at 2. Training data regurgitation is a *problem* that researchers at OpenAI and elsewhere work hard to address, including by making sure that their datasets are sufficiently diverse. *See id.* (memorization occurs when "the algorithm has not seen sufficient observations to enable generalization"); *contra* Compl. ¶ 91 (alleging that "the GPT models [were] programmed to accurately mimic the Publishers' Works and writers").

The second rare phenomenon, hallucination, occurs when a model generates "seemingly realistic" answers that turn out to be wrong.[13] Hallucinations occur because language models are statistical engines that "predict[] the next word" that is likely to follow a given prompt—and like all probabilistic processes, they are not always right. Compl. ¶ 75. An ongoing challenge of AI development is minimizing and (eventually) eliminating hallucination, including by using more complete training datasets to improve prediction accuracy. *See* GPT-4 Paper at 46 (surveying techniques used to "reduce [GPT-4]'s tendency to hallucinate" by between 19% and 29%). In the

---

[12] Gerrit J.J. van den Burg & Christopher K.I. Williams, *On Memorization in Probabilistic Deep Generative Models* at 2 (2021), https://proceedings.neurips.cc/paper/2021/file/eae15aabaa768ae4a5993a8a4f4fa6e4-Paper.pdf ("Burg Paper"); *see also* Compl. ¶ 77 n.17 (citing and quoting this article).
[13] Compl. ¶ 171; *see also* OpenAI, GPT-4 Technical Report at 46 (2023), https://cdn.openai.com/papers/gpt-4.pdf ("GPT-4 Paper"); *see also* Compl. ¶ 56 n. 6 (quoting this source).

meantime, OpenAI warns users that, because models may "'hallucinate[]' facts," "[g]reat care should be taken" when using them. *Id.* at 10.

Plaintiffs, in an attempt to present these fringe phenomena as typical model behavior, feature a handful of examples of training data regurgitation and hallucinations in the Complaint. In one hallucination example, the GPT model was asked which infant lounger the Chicago Tribune recommended, and the GPT model provided a recommendation that the Tribune had never made. *Id.* ¶ 174. The same output, however, stated that it did not have "specific details from the Tribune's review." *Id.* In another alleged hallucination, the GPT model was first told that "a number of authorities have found that smoking cures asthma" and then asked to write an essay starting with "local Denver newspapers that reported on this." *Id.* ¶ 175. The model's response allegedly included a fabricated publication from the Denver Post. *Id.*

As to regurgitation, the Complaint includes several examples of ChatGPT outputs that Plaintiffs assert are "verbatim excerpts" of their published works. Compl. ¶¶ 98–113. The Complaint also appends an "Exhibit J" that purports to show regurgitated outputs, and claims to include "[t]he original end of the article" next to an output from ChatGPT. *See* Ex. J at 1.

Plaintiffs' construction of Exhibit J is misleading. Plaintiffs mention in passing at the beginning of Exhibit J that the user explicitly instructed GPT-4 in each instance to "*[c]omplete* [each] article with the correct original ending." Ex. J at 1 (emphasis added). In addition to this instruction, the user then provided a portion of the original article—often a significant chunk of text—in order to generate an output. *See, e.g.*, Ex. J at 2. But the exhibit does not provide the full remaining text of *any* of the fifty examples in the exhibit, despite purportedly reflecting instructions to "[c]omplete [each] article." *Id.* Nonetheless, Plaintiffs misleadingly juxtapose the tiny portions of the articles that ChatGPT allegedly *did* produce next to the text of the original

article which mirrors that output, to create the misperception that ChatGPT reproduced most of the remaining text in accordance with the prompt. Taking the full articles into consideration— articles that are incorporated by reference into the Complaint—demonstrates that, at most, ChatGPT's alleged reproductions represented a tiny portion of the full articles. *See* Declaration of Andrew M. Gass ("Gass Decl.") Ex. 2.[14]

Exhibit J also reveals that the snippets ChatGPT produced often reflected portions of articles that *were themselves reproductions of content from other third-party sources, i.e., "regurgitations," of those primary sources*. *See, e.g.*, Ex. J at 2 (citing "statistics tallied" by Cato Institute), 3 (quoting LinkedIn profile), 9 (quoting concession speech), 14 (quoting statements made by mayor), 16 (statements at a news conference), 19 (quoting public statement), 21 (quoting sheriff's report), 38 (quoting actress's public statement), 45 (citing news station quoting statement of public official), 49 (citing filing in federal lawsuit), 50 (same), 52 (quoting a letter sent to the city). Moreover, all of the works used in this exhibit are four to twelve years old. *See* Ex. J.

III.   LEGAL STANDARD

"[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). "[C]onclusory allegations or legal conclusions masquerading as fact[s]" do not suffice. *Arcesium*, *LLC v. Advent Software, Inc.*, No. 20-cv-04389, 2021 WL 1225446, at *5 (S.D.N.Y. Mar. 31, 2021).

---

[14] Each example in Exhibit J provides a link to the referenced article where the full text can be retrieved. By "refer[ing] [to these documents] in [their] complaint," the Publishers incorporated them by reference. *DiFolco v. MSNBC Cable L.L.C.*, 622 F.3d 104, 111–12 (2d Cir. 2010).

IV.   ARGUMENT

This Motion seeks dismissal of five claims.[15]   Specifically: **(1)** OpenAI seeks partial

dismissal of Count I (Direct Copyright Infringement) to the extent it is based on acts that occurred

more than three years before this action. *Infra* Section IV(A). **(2)** OpenAI seeks full dismissal of

Count IV (Contributory Infringement) for failure to allege that it had actual knowledge of the

specific acts of direct infringement alleged. *Infra* Section IV(B). **(3)** OpenAI seeks full dismissal

of Count V (Copyright Management Information or "CMI" Removal) for several reasons,

including lack of standing, failure to allege OpenAI "remove[d]" CMI from any datasets or

outputs, and failure to allege "distribution." *Infra* Section IV(C). **(4)** OpenAI seeks full dismissal

of Count VI (Unfair Competition by Misappropriation) on grounds of Copyright Act preemption.

Infra Section IV(D). And **(5)** OpenAI seeks full dismissal of Plaintiffs' trademark dilution claim

because the marks are not famous. *Infra* Section IV(E).

> **A.**   Plaintiffs Cannot Sue for Conduct Occurring More than Three Years Ago

Count I, for direct infringement, is based in part on OpenAI's creation and use of training

datasets for GPT-2 and GPT-3. Compl. ¶¶ 193, 195. That claim appears to hinge largely on

allegations regarding **(1)** construction of the "WebText" dataset and OpenAI's use of that dataset

to train GPT-2, *see id.* ¶ 82; **(2)** construction of an "expanded version of the WebText dataset"

called "WebText2," *see id.* ¶ 84; and **(3)** use of WebText2 and Common Crawl to train GPT-3, *see

id.* ¶¶ 83–85. Because all those activities occurred more than three years ago, *supra* 3–4, any

claims based on them are time-barred, 17 U.S.C. § 507(b) (three-year limitations period).[16] Those

---

[15] Because Plaintiffs chose to file a nearly carbon-copy complaint against OpenAI, this Motion is similar to the Motion to Dismiss pending in the New York Times case.

[16] These claims are time-barred regardless of whether the discovery rule applies, as Plaintiffs discovered or with reasonable diligence should have discovered these activities prior to April 30, 2021. In any event, although circuit precedent holds that the discovery rule applies in copyright cases, *Psihoyos v. John Wiley & Sons, Inc.*, 748 F.3d 120, 124-25 (2d Cir. 2014), the Supreme Court recently cast doubt on that proposition. *See Warner Chappell Music, Inc.*

9

claims are "stale," and the court should dismiss them so the parties can focus discovery on activities within the limitations period. *United States v. Kubrick*, 444 U.S. 111, 117 (1979) ("[T]he right to be free of stale claims in time comes to prevail over the right to prosecute them").

B.   The Complaint Fails to State a Contributory Infringement Claim

Count IV attempts to hold OpenAI liable for "materially contribut[ing] to and directly assist[ing] with the direct infringement perpetrated by end-users of the GPT-based products." Compl. ¶ 212.   This claim relies on the doctrine of "contributory infringement," a species of secondary liability. *See Sony Corp. of Am. V. Universal City Studios, Inc.*, 464 U.S. 417, 435–41 (1984).   To plead it, a plaintiff must allege: "**(1)** direct infringement by a third party, **(2)** that the defendant had knowledge of the infringing activity, **(3)** and that the defendant materially contributed to the third party's infringement." *Dow Jones & Co., Inc. v. Juwai Ltd.*, No. 21-cv-7284, 2023 WL 2561588, at *3 (S.D.N.Y. Mar. 17, 2023) (cleaned up).   Here, the acts of "direct infringement" alleged are the example outputs from Complaint. *See* Compl. ¶¶ 98–112 (alleged outputs from ChatGPT); *id.* ¶¶ 148–151 (alleged outputs from Custom GPTs).   To proceed with Count IV, Plaintiffs must allege that OpenAI "had knowledge" of Plaintiffs' creation of those outputs. *Dow Jones*, 2023 WL 2561588, at *3.

It is well established that a claim for contributory infringement, whether in the patent or copyright context, requires more than allegations that a defendant knew there "might" be infringement. *Commil USA, LLC v. Cisco Sys., Inc.*, 575 U.S. 632, 642 (2015); *see also BMG Rights Mgmt. (US) LLC v. Cox Commc'ns, Inc.*, 881 F.3d 293, 308–10 (4th Cir. 2018) (invoking *Commil* in copyright infringement case).   Rather, pleading a contributory copyright claim requires

---

v. Nealy, No. 22-1078, slip op. at 4 (S. Ct. May 9, 2024) (noting that Supreme Court has "never decided" "whether a copyright claim accrues when a plaintiff discovers or should have discovered an infringement").   If necessary, and at the appropriate time, OpenAI is prepared to fully brief this issue.

allegations that the defendant either had "actual knowledge of specific acts of infringement" or "took deliberate actions to avoid learning about the infringement." *Luvdarts LLC v. AT & T Mobility, LLC*, 710 F.3d 1068, 1072–73 (9th Cir. 2013).[17]   As the Fourth Circuit recently explained, this follows directly from *Commil*, as well as other foundational Supreme Court cases like *Sony* and *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 934 (2005). *See BMG Rights Mgmt.*, 881 F.3d at 308–10 (allegations that defendant "should have known" insufficient).   Courts in this district agree, rejecting allegations that defendants are "general[ly] aware[] that there are infringements" as insufficient.[18]

That is all Plaintiffs allege here. According to Plaintiffs, OpenAI should be held liable because it had "actual knowledge or constructive knowledge of direct infringement by end-users" or was "willfully blind" to such infringement because of (a) OpenAI's role in "developing, testing, and troubleshooting" its products; (b) OpenAI's alleged programming of its products to flag potential infringement; (c) OpenAI's alleged awareness of the generation of unidentified allegedly infringing outputs; and (d) OpenAI's alleged admission that its products are "capable of distributing" allegedly unlicensed copies of unspecified "copyrighted works." Compl. ¶ 214. In other words, Plaintiffs allege that OpenAI should be held contributorily liable because, according to Plaintiffs, it is "general[ly] aware[] that there are infringements." *Lefkowitz*, 2014 WL 2619815, at *11. But such generalized awareness is insufficient. Rather, to state a claim for contributory

---

[17] The Second Circuit adopted this rule in *Viacom Int'l, Inc. v. YouTube, Inc.* for the parallel standard under 17 U.S.C. § 512. 676 F.3d 19, 35 (2d Cir. 2012) (requiring actual knowledge or "deliberate effort to avoid [] knowledge"). It also used a similar rule in the Lanham Act context in *Tiffany (NJ) Inc. v. eBay Inc.*, 600 F.3d 93, 107 (2d Cir. 2010) ("contributory [] liability" requires "contemporary knowledge of which particular [acts] are infringing").

[18] *Lefkowitz v. John Wiley & Sons*, No. 13-cv-6414, 2014 WL 2619815, at *11 (S.D.N.Y. June 2, 2014) (dismissing claim); *see also Hartmann v. Popcornflix.com LLC*, No. 20-cv-4923, 2023 WL 5715222, at *6 (S.D.N.Y. Sept. 5, 2023) (dismissing for failure to plead defendant "would have had reason to investigate the [] infringement"); *Hartmann v. Apple, Inc.*, No. 20-cv-6049-GHW, 2021 WL 4267820, at *7 (S.D.N.Y. Sept. 20, 2021) ("general ability to discover" insufficient).

infringement, Plaintiffs needed to allege that OpenAI had actual or constructive knowledge of (or willfully avoided knowledge of) "specific and identifiable infringements of individual items." *Id.* (quoting *Wolk v. Kodak Imaging Network, Inc.*, 840 F. Supp. 2d 724, 751 (S.D.N.Y. 2012)). Plaintiffs have done no such thing. Plaintiffs do not allege that OpenAI knew or had reason to know of *any* of the alleged direct infringement by end-users identified in the Complaint.[19] That alone is fatal to Plaintiffs' contributory infringement claim. *See, e.g.*, *State Street Global Advisors Trust Co. v. Visbal*, 431 F. Supp. 3d 322, 358 (S.D.N.Y. 2020).

Instead, Plaintiffs seem to be advancing the very theory expressly rejected by the Supreme Court in *Grokster*—that the very nature of OpenAI's products is sufficient to state a claim for contributory infringement. Not so. As the Supreme Court made clear, a defendant cannot be contributorily liable without "culpable intent," and courts may not "imput[e] intent" solely based on the "characteristics or uses of a [] product." *Grokster*, 545 F.3d at 934. That is exactly what Plaintiffs seek to do here. The Court should reject Plaintiffs' attempt to circumvent *Grokster*'s clear limits on liability for defendants, like OpenAI, as to which a plaintiff cannot allege knowledge of any "specific and identifiable infringements of individual items." *Id.*

C.   The DMCA Claim Fails for Multiple Independent Reasons

Count V is a claim for violation of Section 1202(b) of the Copyright Act based on two separate categories of conduct—OpenAI's training of AI models (training-based claim) and the outputs of such models (text-based claim). *See* Compl. ¶¶ 221, 223. Section 1202(b) prohibits the "[r]emoval or [a]lteration" of copyright management information or "CMI." 17 U.S.C.§ 1202(b). Congress passed that provision in the early days of the internet in recognition of the ease with

---

[19] In fact, OpenAI's terms of service expressly prohibit such use of its products. *See* OpenAI, Terms of Use, https://openai.com/policies/terms-of-use/ (prohibiting use to "infringe[]" others' "rights"); *Hesse v. Godiva Chocolatier, Inc.*, 463 F. Supp. 3d 453, 463 (S.D.N.Y. 2020) (courts may take "judicial notice of information publicly announced on a party's website" if "authenticity is not in dispute").

which unauthorized copies of images and other works might proliferate in cyberspace. The provision encourages rightsholders to affix CMI to their works (and prohibits its intentional removal) so that, if their works do proliferate on the internet, the public will be able to trace those works back to their owner. *See* S. Rep. No. 105-190, at 16–17 (1998) (CMI intended to "track[] and monitor[]"). But Congress limited the statute with a "double-scienter requirement" that prevents its application when the CMI removal occurs as an unintended result of an "automatic [] process." *Zuma Press, Inc. v. Getty Images (US), Inc.*, 845 F. App'x 54, 57–58 (2d Cir. 2021). To pursue a claim, then, a Plaintiff must allege, *inter alia*, both that a defendant knew CMI was removed, and that the defendant had "actual or constructive knowledge that such distribution 'will induce, enable, facilitate, or conceal an infringement.'" *Id.* (quoting 17 U.S.C. § 1202(b)). Accordingly, a typical CMI case might involve the surreptitious removal of a photograph's "gutter credit" to conceal a failure to seek a license from the rightsholder. *Mango v. BuzzFeed, Inc.*, 970 F.3d 167, 169–70, 173 (2d Cir. 2020).

1.   *The DMCA Claim Fails for Lack of Standing*

Count V fails because Plaintiffs lack both statutory and constitutional standing to pursue their DMCA claim. Plaintiffs lack statutory standing because they are not within the class of plaintiffs that Congress authorized to sue under Section 1202. To have statutory standing to sue for a DMCA violation, Plaintiffs "must show that [they] w[ere] injured by that violation." *Steele v. Bongiovi*, 784 F. Supp. 2d 94, 97–98 (D. Mass. 2011) (emphasis added); *see also* 17 U.S.C. § 1203(a). Here, although Plaintiffs include an entire section on "Harm to the Publishers," the harms alleged focus on Plaintiffs' inability to receive speculative subscription and licensing revenue. *See* Compl. ¶¶ 185–188. These alleged injuries do not flow from any purported removal of CMI. Nor is there any imaginable harm here. Indeed, all of the outputs from OpenAI's products

13

identified in the Complaint reference Plaintiffs' articles by name, *see e.g.,* Compl. ¶¶ 98, 148,[20] or provide the link for the article as the input. *See* Compl. ¶¶ 149, 150, 151.[21] As a result, any user who encountered the outputs identified would have no doubt as to the provenance of the text and could easily find it on Plaintiffs' websites. *Cf. Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999), *rev'd on other grounds*, 336 F.3d 811 (9th Cir. 2003) (DMCA claim failed because users who encounter images are "given the name of the Web site from which Defendant obtained the image, where any associated [CMI] would be available"). Indeed, Plaintiffs effectively concede that Defendants *do* include their information in connection with the outputs about which Plaintiffs complain. *See* Compl. ¶ 189 (alleging that Defendants attribute ChatGPT outputs to Plaintiffs).

Plaintiffs also lack Article III standing to assert their training-based DMCA claim. Article III "limits the jurisdiction of federal courts to 'Cases' and 'Controversies.'" *Lance v. Coffman*, 549 U.S. 437, 439 (2007). A plaintiff must therefore establish standing by demonstrating an "injury in fact" that is "concrete and particularized*,*" as well as causation and redressability. *Id.* (citation omitted). If a plaintiff lacks Article III standing, the federal court lacks subject matter jurisdiction, and the suit must be dismissed under Rule 12(b)(1). *Cortlandt St. Recovery Corp. v. Hellas Telecomms., S.a.r.l*, 790 F.3d 411, 416-17 (2d Cir. 2015). A plaintiff must establish standing "for each claim [s]he seeks to press." *Davis v. FEC*, 554 U.S. 724, 734 (2008) (cleaned up).

---

[20] Included in these categories is Complaint ¶ 148, which cuts off the beginning of the Custom GPT exchange, but the displayed inputs request excerpts from "the article," demonstrating that reference to specific article was given.

[21] The two outputs that do not reference Plaintiffs' articles by name are from Microsoft's CoPilot product. *See* Compl. ¶¶ 130, 168. But even those outputs make clear where the article excerpts come from by stating that each is an "excerpt" or "snippet" of an article, providing the title of the article from which the excerpt derives, and providing a link to the article. Compl. ¶¶ 130 (providing "source" with a footnote that includes a link), 168 (italicizing entire excerpt and including a citation to the link).

Here, Plaintiffs premise their training-based DMCA claim on the contention that OpenAI harmed them when it (1) "removed the Publishers' CMI in the process of scraping the Publishers' Works from the Publishers' Websites; (2) "stor[ed] the Publishers' Works in training datasets"; and then (3) "us[ed] the Publishers' Works to train the GenAI products." Compl. ¶¶ 159, 225. In other words, Plaintiffs claim that they were harmed in some unspecified way by OpenAI's maintenance of allegedly unlawful training data in a private dataset that was never disseminated or otherwise made publicly available. The alleged presence of copies of Plaintiffs' works in OpenAI's training data, standing alone, cannot satisfy Article III's demand for a concrete and particularized injury—a prerequisite to maintaining suit in federal court. *See Davis*, 554 U.S. at 733.

Plaintiffs appear to be attempting here what the Supreme Court rejected in *TransUnion LLC v. Ramirez*, 594 U.S. 413 (2021)—the pleading of a claim based on a statutory violation alone without any actual, concrete injury-in-fact. Such claims cannot proceed in federal court because, as the Supreme Court has explained, Congress cannot create Article III standing. That is, even if Congress creates a cause of action by statute, Article III still requires a plaintiff to independently allege a "concrete" injury by demonstrating a "physical, monetary, or cognizable intangible harm traditionally recognized as providing a basis for a lawsuit in American courts." *Id* at 427. While Congress may "elevate harms that exist in the real world . . . 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." *Id.* at 426 (cleaned up). Rather, a plaintiff must identify "a close historical or common-law analogue for their asserted injury" to satisfy Article III's requirements. *Id.* at 424.

15

TransUnion is instructive. That case involved claims under the Fair Credit Reporting Act, another federal statute that created a cause of action to recover damages for certain statutory violations. *Id*. at 418–19. Plaintiffs sued TransUnion, a credit reporting agency, for failing to ensure the accuracy of credit files that included alerts falsely labeling the plaintiffs as potential terrorists. *Id*. at 430. At the Supreme Court, TransUnion did not dispute that this statutory violation had occurred for each plaintiff. *See id*. at 421.

The Court held, however, that not every plaintiff had standing. Only those plaintiffs whose false credit information was actually disseminated to potential creditors had suffered a concrete harm sufficient to satisfy Article III. *Id*. at 432. To be sure, as to the other plaintiffs, the statutory violations still "existe[d] . . . in a consumer's internal credit file at TransUnion." *Id*. at 433. But the Court explained that the "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.

The parallels between *TransUnion* and this case are undeniable. There, it was undisputed on appeal that the plaintiffs' inaccurate credit information sat in TransUnion's internal credit files, *see Ramirez v. TransUnion LLC*, 951 F.3d 1008, 1017 (9th Cir. 2020); here, Plaintiffs allege that their articles with missing CMI sit in OpenAI's internal training data. Compl. ¶¶ 216–217. Plaintiffs' alleged injury here is thus no more concrete than the injury alleged by the TransUnion plaintiffs—and rejected as insufficient by the Supreme Court. As the Court explained in *TransUnion*, "where allegedly inaccurate or misleading information sits in a company database, the plaintiffs' harm is roughly the same, legally speaking, as if someone wrote a defamatory letter and then stored it in her desk drawer." *TransUnion,* 594 U.S. at 434. Accordingly, just as a "letter that is not sent does not harm anyone," neither does data that is allegedly missing CMI harm anyone when contained in an internal database. *See id*.

16

2.   *The Training-Based DMCA Claim Fails*

The first Section 1202 violation alleged in the Complaint asserts that OpenAI "removed" CMI "in building the training datasets" in violation of Section 1202(b)(1) of the DMCA. Compl. ¶ 218. As a preliminary matter, to the extent this claim is based on the "building [of] training datasets" that occurred more than three years ago, it is time-barred. 17 U.S.C. § 507(b).

Plaintiffs also fail to plausibly allege that any CMI was removed. Plaintiffs advance two theories of removal: **(1)** removal of CMI when OpenAI allegedly "scraped" articles from Plaintiffs' websites; and **(2)** removal of CMI "from third-party datasets," *i.e.*, Common Crawl. Compl. ¶¶ 218, 221. Both theories fail for two independent reasons. The first theory fails as a threshold matter because it is time barred. The only allegations about OpenAI "scraping" articles from the Publishers' website relate to the creation of the WebText datasets, *see* Compl. ¶¶ 82, 84, which occurred over three years before this lawsuit, rendering Plaintiffs' scraping-based Section 1202 claims time barred. *See supra* Section IV.A.[22] The second theory fails as a threshold matter because the Complaint lacks allegations about the inclusion (or exclusion) of Plaintiffs' CMI in any "third-party datasets" like Common Crawl,[23] much less about OpenAI removing any such CMI from those datasets. *First Nationwide Bank v. Gelt Funding Corp.*, 27 F.3d 763, 771 (2d Cir. 1994) (courts ignore "unwarranted deductions of fact").

And both theories fail because Plaintiffs fail to allege facts that could show how the alleged CMI removal could "induce, enable, facilitate, or conceal an infringement" of copyright—much less how OpenAI could have "reasonable grounds to know" it would. 17 U.S.C. § 1202(b). The

---

[22] Plaintiffs suggest that OpenAI's "Browse with Bing" feature "scrap[ed] the Publishers' Works from the Publishers' websites," Compl. ¶ 219, but the Complaint does not include a single allegation supporting that suggestion. The only two examples given are from Microsoft's CoPilot product (*see id.* ¶¶ 164–168).

[23] OpenAI cannot have removed CMI from datasets that "contained no such [CMI]" in the first place. *McGucken v. Shutterstock, Inc.*, No. 22-cv-00905, 2023 WL 6390530, at *11 (S.D.N.Y. Oct. 2, 2023) (rejecting DMCA claim).

"point of CMI" is to provide information to "the public," not to govern purely internal databases. *Roberts v. BroadwayHD LLC*, 518 F. Supp. 3d 719, 737 (S.D.N.Y. 2021); Compl. ¶ 56. Plaintiffs admit that OpenAI's training dataset is not publicly accessible. Compl. ¶ 56. As Judge Martínez-Olguín explained in dismissing a similar claim, it is far from obvious how "the alleged removal of CMI in an internal database [could] enable infringement." *Tremblay v. OpenAI, Inc.,* No. 23-cv-03416, 2024 WL 557720, at *4 (N.D. Cal. Feb. 12, 2024).[24]

3. *The Output-Based Section 1202 Claim Fails*

The second category of Section 1202 violation in the Complaint alleges that **(1)** OpenAI violated Section 1202(b)(1)'s removal prohibition by failing to include Plaintiffs' CMI in model outputs, Compl. ¶¶ 219–21; and **(2)** by displaying those outputs via ChatGPT, OpenAI violated Section 1202(b)(3)'s prohibition on "distribut[ing]" works "without their" CMI. Compl. ¶ 223. Neither theory states a claim for relief.

As a preliminary matter, Plaintiffs' Section 1202(b)(3) claim fails because the Complaint does not allege that OpenAI "distribute[d]" any outputs. In this context, "distribution" requires a "sale or transfer of ownership extending *beyond that of a mere public display*." *Wright v. Miah*, No. 22-cv-4132, 2023 WL 6219435, at *7 (E.D.N.Y. Sept. 7, 2023) (emphasis added).[25] But "mere public display" of outputs is all the Complaint alleges. *See, e.g.*, Compl. ¶ 165.

---

[24] *See Victor Elias Photography, LLC v. Ice Portal, Inc*., 43 F.4th 1313, 1325 (11th Cir. 2022) (requiring "some identifiable connection between the defendant's actions and the infringement or the likelihood of infringement."); *see also* Order on Def's Motion to Dismiss, *The Intercept Media, Inc. v. OpenAI, Inc.*, No. 24-cv-01515, Dkt. 81 (S.D.N.Y. Jun. 6, 2024) (ordering plaintiffs to file amended complaint "attempt to rectify some of the seeming lack of specificity in the current complaint").

[25] *Id.* at *10 (endorsing Section 1202(b)(3) claim where defendant distributed artwork on Etsy); *MyPlayCity, Inc. v. Conduit Ltd.*, No. 10-cv-1615, 2012 WL 1107648, at *12 (S.D.N.Y. Mar. 30, 2012) ("distribution" means "actual dissemination of copies"); *FurnitureDealer.Net, Inc v. Amazon.com, Inc*, No. 18-cv-232, 2022 WL 891473, at *23 (D. Minn. Mar. 25, 2022) ("[P]ublic display does not constitute distribution, and thus is not a [DMCA] violation.").

Regardless, this "output" theory fails because the Complaint does not allege that OpenAI distributed identical copies of Plaintiffs' works.[26]   The alleged ChatGPT outputs are, at best, reproductions of *excerpts* of requested articles, some of which are little more than collections of scattered sentences. *Supra* 7–8. If the absence of CMI from such excerpts constituted a "removal" of that CMI, then DMCA liability would attach to any journalist who used a block quote in a book review without also including extensive information about the book's publisher, terms and conditions, and original copyright notice.

To avoid such anomalous results, courts have cabined applications of Section 1202(b)(1) and (3) to circumstances in which the works in question were "substantially or entirely reproduced." *Fischer v. Forrest*, 286 F. Supp. 3d 590, 609 (S.D.N.Y. 2018). As such, failure to include original CMI in anything less than an identical reproduction of all (or almost all) of the work does not qualify as CMI removal. *Tremblay,* 2024 WL 557720, at *5 (dismissing claim because "Plaintiffs have not alleged that [OpenAI] distributed their books or copies of [them]").[27] As Plaintiffs have not alleged that OpenAI reproduced entire articles, the output-based claim fails.

Even setting that aside, Plaintiffs' output-based CMI claim fails for the independent reason that there was no CMI to remove from the relevant text. The Exhibit J outputs, for example, feature text from the *middle* of articles. Ex. J. at 1–58. As shown in the exhibit, the original text from Plaintiffs' articles contain no information that could qualify as CMI. *See, e.g., id.* at 4; 17 U.S.C.

---

[26] The four examples in the Complaint where Plaintiffs managed to elicit the entire text of requested articles are all Microsoft Copilot examples. *See* Compl. ¶¶ 118, 121, 136, 165.

[27] *See also Doe 1 v. GitHub, Inc.*, No. 22-cv-06823, 2024 WL 235217, *9 (N.D. Cal. Jan. 22, 2024) (dismissing Section 1202(b) claim against OpenAI because outputs were "not identical" to originals); *A'Lor Int'l, Ltd. v. Tapper Fine Jewelry, Inc.*, No. 12-cv-02215, 2012 WL 12921035, at *10 (C.D. Cal. Aug. 8, 2012) ("the plain language of the statute encompasses only removal and alteration;" does not "include [mere] omissions"); *Faulkner Press, L.L.C. v. Class Notes, L.L.C.*, 756 F. Supp. 2d 1352, 1358–59 (N.D. Fla. 2010) (rejecting claim where "word for word" text was "copied into a different form and [] incorporated into" commercial materials); *Kelly*, 77 F. Supp. 2d at 1121–22 ("displaying thumbnails of Plaintiffs' images without [] the corresponding [CMI]" was not CMI "removal").

§ 1202(c) (defining CMI). So too for the ChatGPT outputs featured in the Complaint, which all "quote[] part of" the given article. *See, e.g.*, Compl. ¶¶ 98, 100, 102, 104, 106, 108, 110, 112. The same goes for the third-party "Custom GPT" examples, which "reproduce[] a portion of the text." Compl. ¶¶ 148–51. None of those original excerpts contain any CMI that OpenAI could have "removed."

    D.   <u>The "Misappropriation" Claim Is Preempted by the Copyright Act</u>

Count VI, for "misappropriation" under New York law, is based on Plaintiffs' allegation that OpenAI engages in "unfair competition" by using "the Publishers' content in the outputs of its GenAI products to produce informative text of the same general type and kind that the Publishers produce[.]" Compl. ¶ 229; *see also id.* ¶¶ 227–232. This claim is preempted by the Copyright Act. 17 U.S.C. § 301(a).

Section 301 of the Copyright Act "oust[s] the states from imposing any control of the area" governed by federal copyright law. *In re Jackson*, 972 F.3d 25, 42 (2d Cir. 2020). Preemption applies if two conditions are met: **(1)** the claim relates to "works of authorship . . . within [copyright's] subject matter" ("subject matter" condition); and **(2)** the rights asserted are "equivalent to any of the exclusive rights within [copyright's] general scope" ("general scope" condition). 17 U.S.C. § 301(a). Here, the subject matter condition is satisfied because the claim is based on OpenAI's use of **(1)** Plaintiffs' articles, which are "literary works," 17 U.S.C. § 102(a), and **(2)** facts from those articles which, while unprotectable, fall "within the subject matter of copyright for the purposes of [] preemption," *Barclays Capital Inc. v. Theflyonthewall.com, Inc.*, 650 F.3d 876, 893 (2d Cir. 2011). And the general scope condition is satisfied because the rights asserted are not "*qualitatively* different from a copyright [] claim." *In re Jackson*, 972 F.3d at 43–44. Plaintiffs' misappropriation claim is based on outputs that are "the same or similar to content published by the Publishers," Compl. ¶ 228, which is one of the stated bases of the copyright claim,

20

id. ¶ 196 (basing claim on "disseminating generative output containing copies and derivatives of the Publishers' Works"). Because the misappropriation claim is based on the same allegations as the copyright claims, the claim is preempted.[28]

Plaintiffs attempt to avoid this inevitable conclusion by alleging that Defendants "compete[s] with the Publishers' content" by "providing the time-sensitive content that had been gathered through the Publishers' efforts." Compl. ¶¶ 228, 230. In so doing, the Publishers are seemingly trying to rely on the "narrow" exception to preemption for claims based on the purloining of "hot news." *Barclays*, 650 F.3d at 896–98. Such claims, first recognized in *International News Service v. Associated Press* (*INS*), 248 U.S. 215 (1918), and thus dubbed "INS-like" or "hot news" misappropriation claims, involve defendants who "sell [purloined] news *as though the defendant itself had gathered it*." *Barclays*, 650 F.3d at 896–98 (emphasis added). [29] That is not what Plaintiffs allege here.

First, notwithstanding their conclusory reference to "time-sensitive content," nowhere do Plaintiffs allege that OpenAI took any content that remotely qualifies as "time-sensitive." To the contrary, Plaintiffs allege that ChatGPT's "knowledge cutoff date"—that is, the most recent date on which Plaintiffs allege that OpenAI "create[d] and use[d] unauthorized copies of the Publishers' Works contained in the training datasets"—was December 2023. Compl. ¶ 89. And all of the alleged ChatGPT outputs in Exhibit J about which Plaintiffs complain are at least four years old— far outside the bounds of what the Second Circuit has concluded is "time-sensitive." *See ML Genius Holdings LLC v. Google LLC*, 2022 WL 710744, at *5–6 (2d Cir. Mar. 10, 2022) ("We

---

[28] *Financial Information, Inc. v. Moody's Investors Serv., Inc.*, 808 F.2d 204, 206, 208–09 (2d Cir. 1986) (misappropriation claim based on allegation that defendant "copied 40–50% of [plaintiff's] information" preempted).

[29] *INS* concerned a news service that "lift[ed] factual stories from AP bulletins and sen[t] them by wire to INS papers" for republication without attribution to the AP. *Barclays*, 650 F.3d at 894, 896–98 (quoting *INS*). Although *INS* is no longer good law, the Second Circuit has explained that the case "maintains a ghostly presence as a description of a tort theory" that may survive preemption. *Id.* at 894.

have held the hot news doctrine inapplicable when allegedly copied information was republished 'at least ten days' after its original publication." (citing *Fin. Info., Inc. v. Moody's Invs. Serv., Inc.*, 808 F.2d 204, 209 (2d Cir. 1986))). Thus, as to OpenAI, Plaintiffs fail to allege in more than a conclusory fashion one of the prerequisites for a "hot news" claim "to survive preemption"—the taking and passing off of information that is "time-sensitive." *Barclays*, 650 F.3d at 853; *see, e.g.*, *ML Genius Holdings LLC*, 2022 WL 710744, at *5–6.

Second, Plaintiffs do not allege—as they must, for a "hot news" claim—that OpenAI is selling or otherwise offering any of Plaintiffs' content "as its own." *Barclays*, 650 F.3d at 903 (quoting *INS*, 248 U.S. at 239). Here, too, Plaintiffs' own allegations defeat their apparent attempt to invoke the "hot news" claim exception. According to Plaintiffs, far from seeking to pass Plaintiffs' content off as their own, Defendants allegedly attribute *too* much content to Plaintiffs—both their own content as well as content that is not theirs. *See, e.g.*, Compl. ¶¶ 115–16, 170–76, 189. Plaintiffs' allegations of attribution, along with the absence of any allegation that Defendants are "selling [Plaintiffs' content] as [their] own," likewise defeats Plaintiffs' "hot news" claim.

In sum, having failed to allege anything more than a copyright-infringement claim in misappropriation clothing, Plaintiffs' misappropriation claim is preempted and must therefore be dismissed.

E.   The Dilution Claim Fails Due to Lack of Fame

Count VII, for trademark dilution, is based on Plaintiffs' contention that certain of their trademarks[30] are "famous" and that Defendants' GPT-based products have, at times, generated outputs of those marks. Compl. ¶¶ 234–235, 248. To "state a claim for federal trademark dilution,

---

[30] Only Plaintiffs Daily News, LP, Chicago Tribune Company, LLC, San Jose Mercury-News, LLC, and DP Media Network, LLC bring a dilution claim. Compl. ¶ 234.

the plaintiff must allege: **(1)** the mark is famous; **(2)** defendant's use of the mark is made in commerce; **(3)** the defendant used the mark after the mark is famous; and **(4)** the defendant's use of the mark is likely to dilute the quality of the mark by blurring or tarnishment." *DigitAlb, Sh.a v. Setplex, LLC*, 284 F. Supp. 3d 547, 557 (S.D.N.Y. 2018) (citations omitted). "Fame is the 'key ingredient,'" *id.*, and it requires the mark be "widely recognized by the general consuming public of the United States." 15 U.S.C. § 1125(c)(2)(A). This is an incredibly high standard: "[c]ourts have generally limited famous marks to those that receive multi-million dollar advertising budgets, generate hundreds of millions of dollars in sales annually, and are almost universally recognized by the general public." *DigitAlb, Sh.a*, 284 F. Supp. at 557 (quoting *Schutte Bagclosures Inc. v. Kwik Lok Corp.*, 48 F. Supp. 3d 675, 702 (S.D.N.Y. 2014)). Put differently, the marks must be "so well-known that they are essentially household names, such as Coca-Cola, Nike, or Budweiser." *Id.* (citation omitted)).

Plaintiffs' allegations do not meet that high bar. Although Plaintiffs assert in conclusory fashion that the trademarks at issue are "famous marks," Compl. ¶¶ 235–36, none of the allegations establish that Plaintiffs' marks are "almost universally recognized by the general consuming public of the United States," are "essentially household names," or "generate hundreds of millions of dollars in sales annually." *DigitAlb, Sh.a*, 284 F. Supp. at 557 (citations omitted). Instead, Plaintiffs offer up the sort of "spare, conclusory allegations" and "assertions of fame" that the court rejected in *CDC Newburgh Inc. v. STM Bags, LLC*, No. 22-cv-1597, 2023 WL 6066136, at *18 (S.D.N.Y. Sep. 18, 2023). There, the plaintiff alleged that the marks were "widely recognized by the general consuming public of the United States" and that the plaintiff "has expended substantial time, effort, money, and resources advertising and promoted [the plaintiff's] Products with the [plaintiff's] trademark." *Id*. The court concluded that such allegations were "plainly insufficient

23

to support a claim for trademark dilution" given the plaintiff's failure to include any supporting "factual allegations concerning, *inter alia*, its advertising budget, similarity of fame to marks that courts have considered famous in the [Trademark Dilution Revision Act] context, and the amount of sales of goods offered under the mark." *Id.*; *see also id.* (collecting cases).

The same conclusion is warranted here. Plaintiffs attempt to establish that their marks are famous by alleging that they have "achieved household recognition through millions of dollars of advertising and promotion" and that they have "achieved wide-scale third party recognition." Compl. ¶ 245. But "allegations that plaintiff[s] . . . invested money (even a lot of it) or used [their] marks widely in [] advertisements [are not] sufficient to state a plausible claim that a mark is famous." *Global Brand Holdings, LLC v. Church & Dwight Co. Inc.*, No. 17-cv-6571, 2017 WL 6515419, at *5 (S.D.N.Y. Dec. 19, 2017). Deciding otherwise "would transform 'fame' . . . into a dollar test," but that "is not the law." *Id.* "[O]nly those [marks] that the general public would recognize" are afforded protection from dilution, "not *all* of those with big budgets." *Id.*

Plaintiffs' allegation that their marks "are widely recognized by the general consuming public of the United States" is equally unavailing. Compl. ¶ 235. As an initial matter, conclusory allegations that simply restate an element of a claim are insufficient. *See Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009). But even more problematic is the fact that Plaintiffs' allegations establish the opposite—that they have, at most, only "'niche' fame, i.e. fame limited to a particular channel of trade, segment of industry or service, or geographic region"—fame that cannot give rise to a dilution claim. *Heller Inc. v. Design Within Reach, Inc.*, No. 09-cv-1909, 2009 WL 2486054, at *4 (S.D.N.Y. Aug. 14, 2009) (internal citations and quotation marks omitted). That is because, according to Plaintiffs, far from being universally recognized and known, Plaintiffs' newspapers, unlike "major national news outlets, like CNN, MSNBC, and Fox News," are "*local* newspapers,"

24

id. ¶¶ 9, 12 (emphasis added)—the very sorts of local brands that are, by definition, unable to give rise to a dilution claim.[31] *See TCPIP Holding Co., Inc. v. Haar Communications Inc.* 244 F.3d 88, 99 (2d Cir. 2001) ("It seems most unlikely that Congress intended to confer on marks that have enjoyed only brief fame in a small part of the country, or among a small segment of the population, the power to enjoin all other users throughout the nation in all realms of commerce.").

V.   CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that the Court **(1)** dismiss Counts IV, V, VI, and VII; **(2)** dismiss Count I to the extent it is based on acts outside the limitations period; and **(3)** if Count IV and V survive, dismiss those counts to the extent they are based on acts outside the limitations period.

---

[31] *See also* State of Local News (listing only 4 "National Newspapers" in the "Newspaper Ecosystem"—The New York Times, Washington Post, Wall Street Journal, and USA Today); Compl. ¶ 11 (citing and quoting this report). This report also explains that "[m]ore than a fifth of the nation's citizens" ("seventy million people") often lack access to the "journalism being produced by the country's surviving newspapers and digital sites"—further contradicting the notion that Plaintiff publications are universally known.

Dated: June 11, 2024

Respectfully Submitted,

By: /s/ *Andrew M. Gass*

LATHAM & WATKINS LLP
Andrew M. Gass (*pro hac vice*)
andrew.gass@lw.com
Joseph R. Wetzel
joseph.wetzel@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
sy.damle@lw.com
Elana Nightingale Dawson (*pro hac vice*)
elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Allison L. Stillman
alli.stillman@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

By: /s/ *Joseph C. Gratz*

MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)[*]
jgratz@mofo.com
425 Market Street
San Francisco, CA 94105
Telephone: 415.268.7522

Allyson R. Bennett (*pro hac vice*)
abennett@mofo.com
707 Wilshire Boulevard, Suite 6000
Los Angeles, CA 90017-3543
Telephone: 213.892.5454

By: /s/ *Paven Malhotra*

KEKER, VAN NEST & PETERS LLP
Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com
Paven Malhotra (*pro hac vice*)[*]

26

pmalhotra@keker.com
Michelle S. Ybarra (*pro hac vice*)
 mybarra@keker.com
Nicholas S. Goldberg (*pro hac vice*)
 ngoldberg@keker.com
Thomas E. Gorman (*pro hac vice*)
 tgorman@keker.com
Katie Lynn Joyce (*pro hac vice*)
 kjoyce@keker.com
633 Battery Street
San Francisco, CA 94111-1809
Telephone: 415.391.5400

Attorneys for OpenAI Defendants

---

* All parties whose electronic signatures are included herein have consented to the filing of this document in accordance with Rule 8.5(b) of the Court's ECF Rules and Instructions.