

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

THE NEW YORK TIMES COMPANY,

Plaintiff,

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. 1:23-cv-11195 (SHS) (OTW)

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

ORAL ARGUMENT REQUESTED

TABLE OF CONTENTS

I. INTRODUCTION 1

II. BACKGROUND 4

 A. OpenAI’s Pioneering Research..... 4

 B. The Key to Generalist Language Models: Scale..... 6

 C. Reliance on Longstanding Fair Use Principles..... 7

 D. GPT-3.5 and GPT-4..... 8

 E. OpenAI’s API and ChatGPT 9

 F. The Times Files Suit..... 10

 G. The Times Focuses on Two Fringe Behaviors: Regurgitation & Hallucination .. 10

 1. Outputs from Developer Tools 12

 2. ChatGPT Outputs..... 13

III. LEGAL STANDARD..... 14

IV. ARGUMENT..... 14

 A. The Times Cannot Sue for Conduct Occurring More than Three Years Ago 15

 B. The Complaint Fails to State a Contributory Infringement Claim 15

 C. The DMCA Claim Fails for Multiple Independent Reasons 17

 1. The Times Did Not Specify the CMI at Issue..... 17

 2. The Training-Based Section 1202 Claim Fails..... 18

 3. The Output-Based Section 1202 Claim Fails..... 20

 4. The Times Fails to Allege a CMI-Based Injury..... 21

 D. The “Misappropriation” Claim Is Preempted by the Copyright Act 22

 1. The Text Claim Fails 22

 2. The Recommendations Claim Fails 23

V. CONCLUSION..... 25

TABLE OF AUTHORITIES

Page(s)

CASES

A’Lor Int’l, Ltd. v. Tapper Fine Jewelry, Inc.,
 No. 12-cv-02215, 2012 WL 12921035 (C.D. Cal. Aug. 8, 2012)21

Andersen v. Stability AI Ltd.,
 No. 23-cv-00201, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023)18

Arcesium, LLC v. Advent Software, Inc.,
 No. 20-cv-04389, 2021 WL 1225446 (S.D.N.Y. Mar. 31, 2021).....14

Ashcroft v. Iqbal,
 556 U.S. 662 (2009).....14

Authors Guild v. Google, Inc. (Google Books),
 804 F.3d 202 (2d Cir. 2015).....3, 8

Authors Guild, Inc. v. HathiTrust,
 755 F.3d 87 (2d Cir. 2014).....3

Baker v. Selden,
 101 U.S. 99 (1879).....3

Barclays Capital Inc. v. Theflyonthewall.com, Inc.,
 650 F.3d 876 (2d Cir. 2011).....23, 24, 25

BMG Rights Mgmt. (US) LLC v. Cox Commc’ns, Inc.,
 881 F.3d 293 (4th Cir. 2018)16

Clanton v. UMG Recordings, Inc.,
 556 F. Supp. 3d 322 (S.D.N.Y. 2021).....3

Commil USA, LLC v. Cisco Sys., Inc.,
 575 U.S. 632 (2015).....16

Design Pics Inc. v. PBH Network, Inc.,
 No. 20-cv-1096, 2020 WL 8413512 (E.D.N.Y. Oct. 27, 2020)18

DiFolco v. MSNBC Cable L.L.C.,
 622 F.3d 104 (2d Cir. 2010).....4

Doe 1 v. GitHub, Inc.,
 No. 22-cv-06823, 2024 WL 235217 (N.D. Cal. Jan. 22, 2024).....21

Dow Jones & Co., Inc. v. Juwai Ltd.,
 No. 21-cv-7284, 2023 WL 2561588 (S.D.N.Y. Mar. 17, 2023).....15, 16

Faulkner Press, L.L.C. v. Class Notes, L.L.C.,
756 F. Supp. 2d 1352 (N.D. Fla. 2010).....21

Financial Information, Inc. v. Moody’s Investors Serv., Inc.,
808 F.2d 204 (2d Cir. 1986).....23, 25

First Nationwide Bank v. Gelt Funding Corp.,
27 F.3d 763 (2d Cir. 1994).....19

Fischer v. Forrest,
286 F. Supp. 3d 590 (S.D.N.Y. 2018).....21

FurnitureDealer.Net, Inc v. Amazon.com, Inc,
No. 18-cv-232, 2022 WL 891473 (D. Minn. Mar. 25, 2022)20

GC2 Incorporated v. Int’l Game Tech. PLC,
255 F. Supp. 3d 812 (N.D. Ill. 2017)18

Google LLC v. Oracle Am., Inc.,
141 S. Ct. 1183 (2021).....3, 8

Hartmann v. Apple, Inc.,
No. 20-cv-6049-GHW, 2021 WL 4267820 (S.D.N.Y. Sept. 20, 2021)16

Hartmann v. Popcornflix.com LLC,
No. 20-cv-4923, 2023 WL 5715222 (S.D.N.Y. Sept. 5, 2023)16, 17

Hesse v. Godiva Chocolatier, Inc.,
463 F. Supp. 3d 453 (S.D.N.Y. 2020).....2

Hoehling v. Universal City Studios, Inc.,
618 F.2d 972 (2d Cir. 1980).....3, 10

In re Jackson,
972 F.3d 25 (2d Cir. 2020).....22, 23

In re UBS Auction Rate Sec. Litig.,
No. 08-cv-2967, 2010 WL 2541166 (S.D.N.Y. June 10, 2010)7, 13

Int’l News Serv. v. Associated Press,
248 U.S. 215 (1918).....3, 24, 25

Kelly v. Arriba Soft Corp.,
336 F.3d 811 (9th Cir. 2003)8

Kelly v. Arriba Soft Corp.,
77 F. Supp. 2d 1116 (C.D. Cal. 1999),
rev'd on other grounds by 336 F.3d 811 (9th Cir. 2003)19, 21, 22

Lefkowitz v. John Wiley & Sons,
No. 13-cv-6414, 2014 WL 2619815 (S.D.N.Y. June 2, 2014)16

Luvdarts LLC v. AT & T Mobility, LLC,
710 F.3d 1068 (9th Cir. 2013)16, 17

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

McDonald v. West,
138 F. Supp. 3d 448 (S.D.N.Y. 2015).....10

McGucken v. Shutterstock, Inc.,
No. 22-cv-00905, 2023 WL 6390530 (S.D.N.Y. Oct. 2, 2023).....19

Med. Educ. Dev. Servs. v. Reed Elsevier Grp., PLC,
No. 05-cv-8665, 2008 WL 4449412 (S.D.N.Y. Sep. 30, 2008)3

Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.,
545 U.S. 913 (2005).....16

MyPlayCity, Inc. v. Conduit Ltd.,
No. 10-cv-1615, 2012 WL 1107648 (S.D.N.Y. Mar. 30, 2012).....20

Roberts v. BroadwayHD LLC,
518 F. Supp. 3d 719 (S.D.N.Y. 2021).....20

Sony Corp. of Am. v. Universal City Studios, Inc.,
464 U.S. 417 (1984).....8, 15, 16

State Street Global Advisors Trust Co. v. Visbal,
431 F. Supp. 3d 322 (S.D.N.Y. 2020).....16

Steele v. Bongiovi,
784 F. Supp. 2d 94 (D. Mass. 2011)21

Tiffany (NJ) Inc. v. eBay Inc.,
600 F.3d 93 (2d Cir. 2010).....16

Tremblay v. OpenAI, Inc.
No. 3:23-cv-03416, 2024 WL 557720 (N.D. Cal. Feb. 12, 2024).....19, 20, 21

United States v. Kubrick,
444 U.S. 111 (1979).....15

Viacom Int’l, Inc. v. YouTube, Inc.,
676 F.3d 19 (2d Cir. 2012).....16

Victor Elias Photography, LLC v. Ice Portal, Inc.,
43 F.4th 1313 (11th Cir. 2022)20

Wood v. Observer Holdings, LLC,
No. 20-cv-07878, 2021 WL 2874100 (S.D.N.Y., July 8, 2021).....18

Wright v. Miah,
No. 22-cv-4132, 2023 WL 6219435 (E.D.N.Y. Sept. 7, 2023)20

Zuma Press, Inc. v. Getty Images (US), Inc.,
845 F. App’x 54 (2d Cir. 2021)17

STATUTES

17 U.S.C.
§ 102.....3, 23
§ 301.....22, 23
§ 507.....15, 18
§ 512.....16
§ 1202..... *passim*
§ 1203.....22

35 U.S.C. § 271(b).....15

CONSTITUTIONAL PROVISIONS

U.S. CONST. Article 1, § 8, cl. 8.....10

OTHER AUTHORITIES

Adam Gross, *I Asked ChatGPT AI to Write a Sonnet*, LinkedIn (Mar. 2023),
[https:// www.linkedin.com/posts/grossadam_i-asked-chatgpt-ai-to-write-a-sonnet-in-iambic-activity-040728019300229120-ZBNa/](https://www.linkedin.com/posts/grossadam_i-asked-chatgpt-ai-to-write-a-sonnet-in-iambic-activity-040728019300229120-ZBNa/).....1

Cade Metz, *Meet GPT-3. It Has Learned to Code (and Blog and Argue)*, N.Y. Times (Nov. 24, 2020),
<https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html>7

Edmund White, *In 'The Talented Mr. Ripley,' A Shape-Shifting Protagonist Who's Up to No Good*, N.Y. Times Style Magazine (Mar. 24, 2021), <https://www.nytimes.com/2021/03/24/t-magazine/talented-mr-ripley-patricia-highsmith.html>8

Farhad Manjoo, *How Do You Know a Human Wrote This?*, N.Y. Times (July 29, 2020), <https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai-automation.html>7

Francesca Paris & Larry Buchanan, *35 Ways Real People Are Using A.I. Right Now*, N.Y. Times (Apr. 14, 2023), <https://www.nytimes.com/interactive/2023/04/14/upshot/up-ai-uses.html>2

George Getschow, *The Best American Newspaper Narratives of 2012*, Project Muse, https://muse.jhu.edu/pub/172/edited_volume/chapter/1142918 (last visited Feb. 11, 2024).....13

Gerrit J.J. van den Burg & Christopher K.I. Williams, *On Memorization in Probabilistic Deep Generative Models* (2021), <https://proceedings.neurips.cc/paper/2021/file/eae15aabaa768ae4a5993a8a4f4fa6e4-Paper.pdf>.....10, 11

Goldman Sachs, *Generative AI Could Raise Global GDP by 7%* (Apr. 5, 2023), <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>1

H.R. Rep. No. 94-1476 (1976).....8

Hurubie Meko, *The Precarious, Terrifying Hours After a Woman Was Shoved Into a Train*, N.Y. Times (May 25, 2023), <https://www.nytimes.com/2023/05/25/nyregion/subway-attack-woman-shoved-manhattan.html?smid=url-share>14

Hurubie Meko, *The precarious, terrifying hours after a woman was shoved into a train*, Seattle Times (May 27, 2023), <https://www.seattletimes.com/nation-world/nation/the-precarious-terrifying-hours-after-a-woman-was-shoved-into-a-train/>14

John Branch, *Snow Fall: The Avalanche at Tunnel Creek*, N.Y. Times, <https://www.nytimes.com/projects/2012/snow-fall/index.html#/?part=tunnel-creek> (last accessed Feb. 11, 2024).....18

Louis Hyman, *It's Not the End of Work. It's the End of Boring Work*, N.Y. Times (Apr. 22, 2023), <https://www.nytimes.com/2023/04/22/opinion/jobs-ai-chatgpt.html>1

McKinsey & Company, *The Economic Potential of Generative AI: The Next Productivity Frontier* (June 14, 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>1

Michael S. Schmidt & Steven Lee Myers, *Panama Law Firm's Leaked Files Detail Offshore Accounts Tied to World Leaders*, N.Y. Times (Apr. 3, 2016), <https://www.nytimes.com/2016/04/04/us/politics/leaked-documents-offshore-accounts-putin.html>.....3

Microsoft, *The Impact of Large Language Models on Scientific Discovery: a Preliminary Study using GPT-4* (Dec. 8, 2023), <https://arxiv.org/pdf/2311.07361.pdf>.....1

N.Y. Times, *Ethical Journalism: A Handbook*, <https://www.nytimes.com/editorial-standards/ethical-journalism.html#introductionAndPurpose> (last accessed Feb. 25, 2024).....2

Nicholas Lemann, *The Panama Papers and the Monster Stories of the Future*, New Yorker (Apr. 14, 2016) <https://www.newyorker.com/news/news-desk/the-panama-papers-and-the-monster-stories-of-the-future>.....3

OpenAI Blog, *Introducing ChatGPT Enterprise* (Aug. 28, 2023), <https://openai.com/blog/introducing-chatgpt-enterprise>10

OpenAI Blog, *Introducing ChatGPT Plus* (Feb. 1, 2023), <https://openai.com/blog/chatgpt-plus>.....9

OpenAI, GPT-2 Model Card, Github, https://github.com/openai/gpt-2/blob/master/model_card.md (last updated Nov. 2019).....5, 6

OpenAI, GPT-4 Technical Report (2023), <https://cdn.openai.com/papers/gpt-4.pdf>.....11

OpenAI, *Language Models are Few-Shot Learners* (July 22, 2020), <https://arxiv.org/pdf/2005.14165.pdf>.....4, 5, 6, 7

OpenAI, *Language Models are Unsupervised Multitask Learners* (Feb. 14, 2019), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.....5, 6

OpenAI, Terms of Use, <https://openai.com/policies/mar-2023-terms>.....2

Pete Wells, *As Not Seen on TV*, N.Y. Times (Nov. 13, 2012), <https://www.nytimes.com/2012/11/14/dining/reviews/restaurant-review-guys-american-kitchen-bar-in-times-square.html>.....18

Ran Ronen, *How Generative AI Tools Like ChatGPT Can Revolutionize Web Accessibility*, VentureBeat (July 8, 2023), <https://venturebeat.com/ai/how-generative-ai-tools-like-chatgpt-can-revolutionize-web-accessibility/>..... 1

Raphael Brion, *In Which Guy Fieri Answers Pete Wells' Many Questions*, Eater (Nov. 14, 2012), <https://www.eater.com/2012/11/14/6522571/in-which-guy-fieri-answers-pete-wells-many-questions> 13

S. Rep. No. 105-190 (1998)..... 17

Sébastien Bubeck, et al., *Sparks of Artificial General Intelligence: Early Experiments with GPT-4* (Apr. 13, 2023), <https://arxiv.org/pdf/2303.12712.pdf>..... 4, 9

Sheena Vasani, *Be My Eyes AI Offers GPT-4-Powered Support for Blind Microsoft Customers*, Verge (Nov. 15, 2023), <https://www.theverge.com/2023/11/15/23962709/microsoft-blind-users-open-ai-chatgpt-4-be-my-eyes> 1

Uday Dandavate, *How I Used ChatGPT to Write Haiku*, Medium (Sept. 2, 2023), <https://uday-dandavate.medium.com/how-i-used-chatgpt-to-write-haiku-5904ee96360d>..... 1

Zetolgam, *The Chatbot Limerick Writer*, AllPoetry (Dec. 2022) <https://allpoetry.com/poem/16903411-The-Chatbot-Limerick-Writer-by-Zetolgam> 1

I. INTRODUCTION

The artificial intelligence (AI) tool known as ChatGPT is many things: a revolutionary technology with the potential to augment human capabilities, fostering our own productivity and efficiency;¹ an accelerator for scientific and medical breakthroughs;² a mechanism for making existing technologies accessible to more people;³ an aid to help the visually impaired navigate the world;⁴ a creative tool that can write sonnets, limericks, and haikus;⁵ and a computational engine that reasonable estimates posit may add trillions of dollars of growth across the global economy.⁶

Contrary to the allegations in the Complaint, however, ChatGPT is not in any way a substitute for a subscription to The New York Times. In the real world, people do not use ChatGPT or any other OpenAI product for that purpose. Nor could they. In the ordinary course, one cannot use ChatGPT to serve up Times articles at will.

The Times has sought to paint a different picture. Its lawsuit alleges that OpenAI has imperiled the very enterprise of journalism, illustrating the point with 100 examples in which some version of OpenAI’s GPT-4 model supposedly generated several paragraphs of Times content as outputs in response to user prompts. *See* Dkt. 1-68 (Exhibit J).

¹ Louis Hyman, *It’s Not the End of Work. It’s the End of Boring Work*, N.Y. Times (Apr. 22, 2023), <https://www.nytimes.com/2023/04/22/opinion/jobs-ai-chatgpt.html>.

² Microsoft, *The Impact of Large Language Models on Scientific Discovery: a Preliminary Study Using GPT-4* (Dec. 8, 2023), <https://arxiv.org/pdf/2311.07361.pdf> (deployments in “biology and materials design” and “drug discovery”).

³ Ran Ronen, *How Generative AI Tools Like ChatGPT Can Revolutionize Web Accessibility*, VentureBeat (July 8, 2023), <https://venturebeat.com/ai/how-generative-ai-tools-like-chatgpt-can-revolutionize-web-accessibility/>.

⁴ Sheena Vasani, *Be My Eyes AI Offers GPT-4-Powered Support for Blind Microsoft Customers*, Verge (Nov. 15, 2023), <https://www.theverge.com/2023/11/15/23962709/microsoft-blind-users-open-ai-chatgpt-4-be-my-eyes>.

⁵ Adam Gross, *I Asked ChatGPT AI to Write a Sonnet*, LinkedIn (Mar. 2023), https://www.linkedin.com/posts/grossadam_i-asked-chatgpt-ai-to-write-a-sonnet-in-iambic-activity-040728019300229120-ZBNa/; Zetolgam, *The Chatbot Limerick Writer*, AllPoetry (Dec. 2022) <https://allpoetry.com/poem/16903411-The-Chatbot-Limerick-Writer-by-Zetolgam>; Uday Dandavate, *How I Used ChatGPT to Write Haiku*, Medium (Sept. 2, 2023), <https://uday-dandavate.medium.com/how-i-used-chatgpt-to-write-haiku-5904ee96360d>.

⁶ McKinsey & Company, *The Economic Potential of Generative AI: The Next Productivity Frontier* (June 14, 2023), <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> (“generative AI could add the equivalent of \$2.6 trillion to \$4.4 trillion annually” to the global economy); *see also* Goldman Sachs, *Generative AI Could Raise Global GDP by 7%* (Apr. 5, 2023), <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>.

The allegations in the Times’s Complaint do not meet its famously rigorous journalistic standards.⁷ The truth, which will come out in the course of this case, is that the Times paid someone to hack OpenAI’s products. It took them tens of thousands of attempts to generate the highly anomalous results that make up Exhibit J to the Complaint. They were able to do so only by targeting and exploiting a bug (which OpenAI has committed to addressing) by using deceptive prompts that blatantly violate OpenAI’s terms of use.⁸ And even then, they had to feed the tool portions of the very articles they sought to elicit verbatim passages of, virtually all of which already appear on multiple public websites. Normal people do not use OpenAI’s products in this way.⁹

Journalism has undergone many changes in the digital age, and may well undergo more with the advent of AI. OpenAI has established important partnerships with many leaders in the news industry—from large enterprises like the Associated Press and Axel Springer to the dozens of smaller and local outlets associated with the American Journalism Project—to creatively explore and implement AI solutions that assist investigative reporting, create enhanced reader experiences, and improve business operations. The Times’s suggestion that the contrived attacks of its hired gun show that the Fourth Estate is somehow imperiled by this technology is pure fiction. So too is its implication that the public *en masse* might mimic its agent’s aberrant activity.

There is a genuinely important issue at the heart of this lawsuit—critical not just to OpenAI, but also to countless start-ups and other companies innovating in this space—that is being litigated both here and in over a dozen other cases around the country (including in this Court): whether it

⁷ Cf. N.Y. Times, *Ethical Journalism: A Handbook*, <https://www.nytimes.com/editorial-standards/ethical-journalism.html#introductionAndPurpose> (last accessed Feb. 25, 2024) (“we tell our readers the complete, unvarnished truth as best we can learn it”; “Staff members...may not commit illegal acts of any sort.”).

⁸ OpenAI, Terms of Use § 2(c), <https://openai.com/policies/mar-2023-terms> (prohibiting use to “infringe[]” others’ “rights” or “extract data”); *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”).

⁹ Francesca Paris & Larry Buchanan, *35 Ways Real People Are Using A.I. Right Now*, N.Y. Times (Apr. 14, 2023), <https://www.nytimes.com/interactive/2023/04/14/upshot/up-ai-uses.html> (people use ChatGPT to “[w]rite a [] speech,” “[s]kim dozens of academic articles,” “[a]ppeal an insurance denial,” and “[c]reate new proteins in minutes”).

is fair use under copyright law to use publicly accessible content to train generative AI models to learn about language, grammar, and syntax, and to understand the facts that constitute humans' collective knowledge. OpenAI and the other defendants in these lawsuits will ultimately prevail because no one—not even the New York Times—gets to monopolize facts¹⁰ or the rules of language.¹¹ For good reason, there is a long history of precedent holding that it is perfectly lawful to use copyrighted content as part of a technological process that (as here) results in the creation of new, different, and innovative products.¹² Established copyright doctrine will dictate that the Times cannot prevent AI models from acquiring knowledge about facts, any more than another news organization can prevent the Times itself from re-reporting stories it had no role in investigating.¹³ As Justice Brandeis explained more than 100 years ago: “The general rule of law is, that the noblest of human productions—knowledge, truths ascertained, conceptions, and ideas—become, after voluntary communication to others, free as the air to common use.”¹⁴

All of that said, even assuming (counterfactually) the truth of what the lawsuit alleges, several of the theories in the Complaint are not viable, even as pleaded. This Motion asks the Court to trim those at the outset to focus the litigation on the core issues that really matter. In

¹⁰ See, e.g., *Baker v. Selden*, 101 U.S. 99, 103 (1879) (“The very object of publishing a book ... is to communicate to the world the useful knowledge which it contains. But this object would be frustrated if the knowledge could not be used without incurring guilt of piracy of the book.”); *Hoehling v. Universal City Studios, Inc.*, 618 F.2d 972, 979 (2d Cir. 1980) (“[F]actual information is in the public domain.”).

¹¹ See 17 U.S.C. 102(b); *Clanton v. UMG Recordings, Inc.*, 556 F. Supp. 3d 322, 332 (S.D.N.Y. 2021) (“ordinary building blocks of the English language” not protectable); *Med. Educ. Dev. Servs. v. Reed Elsevier Grp., PLC*, No. 05-cv-8665, 2008 WL 4449412, at *6 (S.D.N.Y. Sep. 30, 2008) (no copyright for “concepts such as rules of punctuation, analogies, vocabulary or other fundamental elements of English composition”).

¹² See, e.g., *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) (fair use to create “digital copies of more than ten million [books]” for search tool); *Authors Guild v. Google, Inc. (Google Books)*, 804 F.3d 202 (2d Cir. 2015) (fair use to scan millions of copyrighted books to create novel tool); *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021) (fair use to replicate copyrighted software programming interfaces to create a new mobile platform).

¹³ See Nicholas Lemann, *The Panama Papers and the Monster Stories of the Future*, New Yorker (Apr. 14, 2016) <https://www.newyorker.com/news/news-desk/the-panama-papers-and-the-monster-stories-of-the-future> (noting the Times's refusal to participate in consortium that broke “Panama Papers” story); Michael S. Schmidt & Steven Lee Myers, *Panama Law Firm's Leaked Files Detail Offshore Accounts Tied to World Leaders*, N.Y. Times (Apr. 3, 2016), <https://www.nytimes.com/2016/04/04/us/politics/leaked-documents-offshore-accounts-putin.html>.

¹⁴ *Int'l News Serv. v. Associated Press*, 248 U.S. 215, 250 (1918) (Brandeis, J., dissenting).

short: **(1)** The direct copyright infringement claim asserts liability in part from conduct that is time-barred because it occurred more than three years ago. **(2)** The contributory infringement claim would ascribe liability to OpenAI based on generalized knowledge of third-party infringement, rather than actual knowledge of specific infringements, which the law requires. **(3)** The claim for violations of 17 U.S.C. § 1202 (the “DMCA”) fails for 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 Times seeks to ascribe liability. And **(4)** the claim for state common law misappropriation is preempted by the federal Copyright Act.

OpenAI respectfully seeks an order dismissing these legally infirm portions of the Complaint, so that the parties can properly and efficiently litigate the balance.

II. BACKGROUND

A. OpenAI’s Pioneering Research

OpenAI was founded in 2015 to “advance digital intelligence in the way that is most likely to benefit humanity as a whole.” Compl. ¶ 56. It entered the field of “natural language processing” (NLP), which includes the development of statistical tools called “language models.”¹⁵ These models can “predict[] words that are likely to follow a given string of text” based on statistics derived from a body of text—much like a weather model can predict the rain using statistics derived from historical weather data. Compl. ¶ 75. By 2015, research had already unlocked “substantial progress” on “tasks such as reading comprehension” and “question answering.”¹⁶

¹⁵ 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”); Compl. ¶¶ 71, 91 nn.9 & 24 (citing articles). By “refer[ing] [to these documents] in [its] complaint,” the Times incorporated them by reference. *DiFolco v. MSNBC Cable L.L.C.*, 622 F.3d 104, 111–12 (2d Cir. 2010).

¹⁶ OpenAI, *Language Models are Few-Shot Learners* at 3 (July 22, 2020), <https://arxiv.org/pdf/2005.14165.pdf> (“GPT-3 Paper”); *see also* Compl. ¶¶ 86, 90 & nn.18, 22 (citing and quoting this paper).

Those early models, however, were “brittle” and “narrow.”¹⁷ Researchers built them by “manually creat[ing] and label[ing]” datasets to “demonstrate[e] correct behavior”—like sets of English-to-French text translations—and using that data to “train a system to imitate [that] behavior[.]” GPT-2 Paper at 1, 3. The resulting models, while impressive, could only carry out the specific tasks demonstrated by the training data. *Id.*; GPT-3 Paper at 3 (“need for task-specific datasets” was “a major limitation”). “To be broadly useful” to ordinary people, language models needed the ability to “seamlessly mix together or switch between many tasks and skills” without being specifically trained to carry out each task. GPT-3 Paper at 4. In other words, the models needed to be “competent generalists,” not “narrow experts.” GPT-2 Paper at 1.

OpenAI’s researchers set out to solve that complex, scientific problem. In 2019, they posited that the way to build more capable, generalist models was to use “as large and diverse a dataset as possible [] to collect natural language demonstrations of tasks in as varied of domains and contexts as possible.” GPT-2 Paper at 3. The hypothesis was that “[]training at a large enough scale [might] offer a ‘natural’ broad distribution of tasks implicitly contained in predicting the text itself.” GPT-3 Paper at 40. So instead of training its models “on a single domain of text,” OpenAI chose to use a richer and more diverse source: the Internet. GPT-2 Paper at 3.

OpenAI’s researchers identified text from webpages whose URLs had been publicly shared on a social media platform. *Id.* This became a dataset called “WebText,” which OpenAI used to train a model called “GPT-2.” *Id.*; *see also* Compl. ¶ 85. WebText contained a wide array of text from internet forums, restaurant reviews, recipe websites, blogs, shopping websites, dictionaries, medical websites, how-to pages, and more.¹⁸ The dataset was so diverse that even though Times

¹⁷ OpenAI, *Language Models are Unsupervised Multitask Learners* at 1 (Feb. 14, 2019), https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf (“GPT-2 Paper”); *see also* Compl. ¶ 85 n.15 (citing and quoting this paper).

¹⁸ *See* OpenAI, GPT-2 Model Card, Github, https://github.com/openai/gpt-2/blob/master/model_card.md (last

content represented only a tiny fraction of the data, the “NYTimes” was one of the “top 15 domains by volume” in the collection. *See* GPT-2 Model Card. This happened not because OpenAI believed Times articles are more “valu[able]” than other content, *contra* Compl. ¶ 2 (suggesting OpenAI intentionally “gave Times content particular emphasis”), but because of the frequency with which certain social media users shared links to the Times’s content, *see* GPT-2 Paper at 3.

The results of this sophisticated research were impressive. The GPT-2 model proved able to answer trivia questions and perform higher-function tasks like “resolv[ing] ambiguities in text.” GPT-2 Paper at 6–7. The model even showed a “surprising” ability to translate French to English, even though OpenAI had “deliberately removed non-English webpages” from the training dataset. *Id.* at 7. These research results were “exciting” not only because of the model’s capability, but because they scientifically confirmed that the ability to “perform commonsense reasoning” increased dramatically with the size and diversity of the training data. *Id.* at 6 (Figure 3).

B. The Key to Generalist Language Models: Scale

So OpenAI’s research took the logical next step: “scaling up [] model size” by increasing both the “size” and “diversity” of the training data. *See* GPT-3 Paper at 6. To build its next generation of models, OpenAI’s researchers gathered a more robust set of data in part by “expand[ing]” WebText database into a new version called “WebText2,” which included material shared over a longer period of time. *Id.* at 8–9. They also used a filtered version of Common Crawl, a repository of data collected by a non-profit research organization representing “a copy of the Internet.” Compl. ¶ 88. OpenAI disclosed all of this no later than July 22, 2020. GPT-3 Paper at 8. At the time, it was common knowledge that WebText2 and Common Crawl included numerous articles published by the Times.¹⁹

updated Nov. 2019) (“GPT-2 Model Card”); *see also* Compl. ¶ 85 nn. 14, 16, 17 (citing and quoting this source).

¹⁹ Compl. ¶ 88 (citing a “2019 snapshot of Common Crawl”); GPT-2 Model Card (noting prevalence of Times content

The result of this simple act of “scaling up” the training data was, as the Times reported at the time, “mind blowing.”²⁰ The new GPT-3 model was “by far the most powerful ‘language model’ ever created.” Manjoo, *supra* note 20. It could conduct “on-the-fly reasoning” and “unscrambl[e] words, perform[] arithmetic, and us[e] novel words in a sentence after seeing them defined only once.” GPT-3 Paper at 5. Increasing the scale of training led to a surprising jump in ability. *See, e.g., id.* at 22 (Figure 3.10). According to the Times’s reporting, GPT-3 showed that “[m]achines are gaining the ability to write.” Manjoo, *supra* note 20. Within days, developers began to use it to build unprecedented tools. *Id.* (“service that responds to email on your behalf”).

The “key advance,” as the Times reported, was “GPT-3’s flexibility.” *Id.* And the reason it was flexible was OpenAI’s decision to “scal[e] up” the “size” of the training data. GPT-3 Paper at 6. The amount of data needed was staggering. Compl. ¶ 85. But it was that “unprecedented scale” that allowed the model to internalize not only a “map of human language,” but achieve a level of adaptability—and “emergent” intelligence—that “no one thought possible.”²¹

C. Reliance on Longstanding Fair Use Principles

By July 2020, OpenAI had disclosed that Times articles were a tiny part of the diverse datasets that had been used to train these language models. And according to the Complaint, by the time GPT-3 was released in mid-2020, OpenAI had already established itself as a “commercial enterprise.” Compl. ¶ 57. The Times itself reported in 2020 that “OpenAI plans to sell access to GPT-3 via the internet, turning it into a widely used commercial product.” Metz, *supra* note 21. While its reporters joked they “might be put out to pasture by a machine,” Manjoo, *supra* note 20,

in WebText); GPT-3 Paper at 8 (noting WebText2 is an “expanded version of the WebText dataset”).

²⁰ Farhad Manjoo, *How Do You Know a Human Wrote This?*, N.Y. Times (July 29, 2020), <https://www.nytimes.com/2020/07/29/opinion/gpt-3-ai-automation.html>. The Court may take judicial notice of this and other news articles cited in this Motion. *In re UBS Auction Rate Sec. Litig.*, No. 08-cv-2967, 2010 WL 2541166, at *10-12 (S.D.N.Y. June 10, 2010) (judicial notice of “several news items” including *Wall Street Journal* articles).

²¹ Cade Metz, *Meet GPT-3. It Has Learned to Code (and Blog and Argue)*, N.Y. Times (Nov. 24, 2020), <https://www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html>.

the Times never accused OpenAI of violating copyright law. Instead, the Times enthusiastically and factually reported that the technology could be “enormously useful” and “open[] the door to a wide range of new possibilities.” Manjoo, *supra* note 20; Metz, *supra* note 21.

Indeed, it has long been clear that the non-consumptive use of copyrighted material (like large language model training) is protected by fair use—a doctrine as important to the Times itself as it is to the American technology industry.²² Since Congress codified that doctrine in 1976, *see* H.R. Rep. No. 94-1476, at 65–66 (1976) (courts should “adapt” defense to “rapid technological change”), courts have used it to protect useful innovations like home video recording, internet search, book search tools, reuse of software APIs, and many others.²³

These precedents reflect the foundational principle that copyright law exists to control the *dissemination* of works in the marketplace—not to grant authors “absolute control” over all uses of their works. *Google Books*, 804 F.3d at 212. Copyright is not a veto right over transformative technologies that leverage existing works *internally*—*i.e.*, without disseminating them—to new and useful ends, thereby furthering copyright’s basic purpose *without* undercutting authors’ ability to sell their works in the marketplace. *See supra* note 23. And it is the “basic purpose” of fair use to “keep [the] copyright monopoly within [these] lawful bounds.” *Oracle*, 141 S. Ct. at 1198. OpenAI and scores of other developers invested billions of dollars, and the efforts of some of the world’s most capable minds, based on these clear and longstanding principles.

D. GPT-3.5 and GPT-4

OpenAI built on that success by researching, developing, and releasing two other models—GPT-3.5 (in 2022) and GPT-4 (in 2023)—which triggered the AI revolution that we are living

²² *See, e.g.*, Edmund White, In ‘The Talented Mr. Ripley,’ A Shape-Shifting Protagonist Who’s Up to No Good, N.Y. Times Style Magazine (Mar. 24, 2021), <https://www.nytimes.com/2021/03/24/t-magazine/talented-mr-ripley-patricia-highsmith.html> (including continuous 200-word excerpt from published novel).

²³ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 454–55 (1984); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818–22 (9th Cir. 2003); *Google Books*, 804 F.3d at 209; *Oracle*, 141 S. Ct. at 1209 (2021).

through today. GPT-4 “can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more.” Bubeck Paper at 1. Researchers already use these models for a growing list of astonishingly productive ends. *See supra* 1.

E. OpenAI’s API and ChatGPT

Today, there are two primary ways to interact with OpenAI’s models. The first is an application programming interface (“API”) that allows developers to prompt the models directly with computer code. Compl. ¶ 62. Developers must specify a number of parameters, including their desired “Model” version and “Temperature” (*i.e.*, the degree of randomness). *See, e.g., id.* ¶ 140. Developers can supply a “System” prompt, which tells the model how to behave, *e.g.*, “You are a thespian who speaks only in iambic pentameter.” *Id.* OpenAI also provides a “Playground” for developers to experiment with the models. *See id.*

The second is ChatGPT, a consumer-friendly platform—accessible for free at chat.openai.com—to “chat” with an OpenAI model through a user interface.²⁴ Its “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. ¶¶ 72, 108, 112. Using it is the equivalent of instructing a research assistant to use the Internet to look up a fact she does not yet know. The feature goes hand-in-hand with others that collectively amplify ChatGPT’s utility. *See* Bubeck Paper at 45 (describing ability to “coordinate a dinner” using “APIs to retrieve information about the user’s calendar, coordinate with other people over email, [and] book the dinner”).

After its release in November 2022, ChatGPT became an “instant viral sensation.” Compl. ¶ 61. According to the Complaint, by April 2023, the service had 173 million users globally who rely on ChatGPT to “craft clearer communications, accelerate coding tasks, rapidly explore

²⁴ Compl. ¶¶ 62, 102; OpenAI Blog, *Introducing ChatGPT Plus* (Feb. 1, 2023), <https://openai.com/blog/chatgpt-plus>; *see also* Compl. ¶ 149 n.41 (citing this page).

answers to complex business questions, assist with creative work, and much more.”²⁵

F. The Times Files Suit

It was only after this rapid adoption, along with reports of the value unlocked by these new technologies, that the Times claimed that OpenAI had “infringed its copyright[s]” and reached out to demand “commercial terms.” Compl. ¶¶ 54, 126. After months of discussions, the Times filed suit two days after Christmas, demanding “billions of dollars.” *Id.* ¶ 9.

In its suit, the Times now claims ChatGPT will be the end of “independent journalism.” *Id.* ¶¶ 2, 47. Its chief complaint is that OpenAI “use[s]” news articles to build a service the world has never before seen that, according to the Times, might “steal [its] audiences.” *Id.* ¶¶ 2, 8. And its core legal theory is that copyright—a law singularly devoted to the “Progress of Science and useful Arts,” U.S. CONST. art. 1, § 8, cl. 8—should protect it from that technological innovation.

G. The Times Focuses on Two Fringe Behaviors: Regurgitation & Hallucination

To support its narrative, the Times claims OpenAI’s tools can “closely summarize[]” the facts it reports in its pages and “mimic[] its expressive style.” Compl. ¶ 4. But the law does not prohibit reusing facts or styles.²⁶ If it did, the Times would owe countless billions to other journalists who “invest[] [] enormous amount[s] of time, money, expertise, and talent” in reporting stories, Compl. ¶ 32, only to have the Times summarize them in its pages, *see supra* note 13.

To avoid that problem, the Times focuses its allegations on two uncommon and unintended phenomena: **(1)** training data regurgitation and **(2)** model hallucination. The first occurs when a language model “generat[es] a sample that closely resembles [its] training data.”²⁷ This most often

²⁵ Compl. ¶ 149; OpenAI Blog, *Introducing ChatGPT Enterprise* (Aug. 28, 2023), <https://openai.com/blog/introducing-chatgpt-enterprise>; *see also* Compl. ¶¶ 63, 149 & nn.5 & 42 (citing this article);

²⁶ *Hoehling*, 618 F.2d at 978 (“[T]here cannot be any such thing as copyright in the order of presentation of the facts, nor, indeed, in their selection.” (quoting Judge Learned Hand)); *McDonald v. West*, 138 F. Supp. 3d 448, 455 (S.D.N.Y. 2015) (“[C]opyright does not protect styles,” and “[f]or the same reason” it “does not protect ideas”).

²⁷ Gerrit J.J. van den Burg & Christopher K.I. Williams, *On Memorization in Probabilistic Deep Generative Models*

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. Put simply, a model trained on the same block of text multiple times will be more likely to complete that text verbatim when prompted to do so—in the same way that any American who hears the words “I pledge allegiance” might reflexively respond with the words “to the flag of the United States of America.” Training data regurgitation—sometimes referred to as unintended “memorization” or “overfitting”—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. ¶ 93 (alleging that “the GPT models [were] programmed to accurately mimic The Times’s content and writers”).

The second phenomenon—hallucination—occurs when a model generates “seemingly realistic” answers that turn out to be wrong.²⁸ Hallucinations occur because language models are *not* databases of information, but statistical engines that “predict[] words that are likely to follow” a given prompt. Compl. ¶ 75. Like all probabilistic processes, they are not always 100% correct. An ongoing challenge of AI development is minimizing and (eventually) eliminating hallucination, including by using more complete training datasets to improve the accuracy of the models’ predictions. *See* GPT-4 Paper at 46 (surveying techniques used to “reduce [GPT-4]’s tendency to hallucinate” by between 19% and 29%). In the meantime, OpenAI warns users that, because models “‘hallucinate’ facts,” “[g]reat care should be taken” when using them. *Id.* at 10.

In an attempt to frame these undesirable phenomena as typical model behavior, the Complaint features a number of examples of training data regurgitation and model hallucination

at 2 (2021), <https://proceedings.neurips.cc/paper/2021/file/ea15aaba768ae4a5993a8a4f4fa6e4-Paper.pdf> (“Burg Paper”); *see also* Compl. ¶ 80 n.10 (citing and quoting this article).

²⁸ Compl. ¶ 137; *see also* OpenAI, GPT-4 Technical Report at 46 (2023), <https://cdn.openai.com/papers/gpt-4.pdf> (“GPT-4 Paper”); *see also* Compl. ¶ 59 n. 3 (quoting this source).

generated by the Times after what appears to have been prolonged and extensive efforts to hack OpenAI’s models. Notably, the Times does not allege in the Complaint that it made any attempt to share these results with OpenAI (despite being asked to do so), or otherwise collaborate with OpenAI’s ongoing efforts to prevent these kinds of outputs. Rather, the Times kept these results to itself, apparently to set up this lawsuit. The Times’s examples fall into two categories: **(1)** outputs generated by OpenAI’s models using its developer tools and **(2)** ChatGPT outputs.

1. Outputs from Developer Tools

Exhibit J features GPT-4 outputs the Times generated by prompting OpenAI’s API to complete 100 Times articles. Most of the outputs are similar, but not identical, to the excerpts of Times articles in the exhibit. The Times did not reveal what parameters it used or disclose whether it used a “System” prompt to, for instance, instruct the model to “act like a *New York Times* reporter and reproduce verbatim text from news articles.” *See supra* 9. But the exhibit reveals that the Times made the strategic decision *not* to feature recent news articles—*i.e.*, articles that Times subscribers are most likely to read on the Times’s website—but to instead feature much older articles published between 2.5 and 12 years before the filing of the Complaint.²⁹

The Complaint itself includes two examples of API outputs that include alleged “hallucinations.” In the first, the Times used the API Playground to request an essay on how “major newspapers” have reported on “Orange [sic] Juice” and “non-hodgkin’s lymphoma,” and ChatGPT generated a response referencing a non-existent Times article. *See* Compl. ¶ 140. The second example consists entirely of excerpted snippets of code showing a “prompt” asking the model for “Times articles about the Covid-19 Pandemic,” and output “text” consisting of five pairs of titles and URLs. *Id.* The Times claims this output “mislead[s] users” and “tarnish[es]” its

²⁹ *See* Ex. J. at 2–5 (articles published in 2012 and 2019), *id.* at 6–126 (articles published in 2020 and 2021).

marks. *Id.* ¶¶ 142, 202. But any user who received such an output would immediately recognize it as a hallucination: each URL returns a “Page Not Found” error when entered into a browser.

2. *ChatGPT Outputs*

ChatGPT. The Complaint includes two examples of ChatGPT allegedly regurgitating training data consisting of Times articles. Compl. ¶¶ 104–07. In both, the Times asked ChatGPT questions about popular Times articles, including by requesting quotes. *See, e.g., id.* ¶ 106 (requesting “opening paragraphs,” then “the next sentence,” then “the next sentence,” etc.). Each time, ChatGPT provided scattered and out-of-order quotes from the articles in question.³⁰

In its Complaint, the Times reordered those outputs (and used ellipses to obscure their original location) to create the false impression that ChatGPT regurgitated sequential and uninterrupted snippets of the articles. *Compare id.* ¶ 107, with *supra* note 30. In any case, the regurgitated text represents only a fraction of the articles, *see, e.g.,* Compl. ¶ 104 (105 words from 16,000+ word article), all of which the public can already access for free on third-party websites.³¹

Browse with Bing. The Complaint also includes two examples of interactions with “Browse with Bing” created using the same methods. Compl. ¶¶ 118–22. In both, ChatGPT returned short snippets of Times articles. *See id.* ¶ 118 (reproducing first two paragraphs before refusing subsequent request for more); *id.* ¶ 121 (reproducing snippets from first, fourth, and fifth paragraphs). The Complaint suggests that ChatGPT obtained this text from third-party websites.³²

³⁰ *See* Compl. ¶ 104 (providing article’s first two sentences in response to request for “first paragraph;” ignoring request for “next paragraph” and instead providing quote beginning with article’s fifth paragraph); *id.* ¶ 106 (in response to request for “opening paragraphs” and four requests for “the next sentence,” providing snippets of text from first, second, 26th, 27th, eighth, and ninth paragraphs, in that order).

³¹ *See, e.g.,* George Getschow, *The Best American Newspaper Narratives of 2012*, Project Muse, https://muse.jhu.edu/pub/172/edited_volume/chapter/1142918 (last visited Feb. 11, 2024); Raphael Brion, *In Which Guy Fieri Answers Pete Wells’ Many Questions*, Eater (Nov. 14, 2012), <https://www.eater.com/2012/11/14/6522571/in-which-guy-fieri-answers-pete-wells-many-questions>; *see UBS Auction Rate*, 2010 WL 2541166, at *10-12, 15 (courts take judicial notice of articles when not used “for the[ir] truth”).

³² *See* Compl. ¶ 121 (ChatGPT linking to “dnyuz.com”). The regurgitated text in paragraph 118 includes a dateline (“NEW YORK”) that does not appear on the Times’s website, but does appear on other third-party sites in which the

Wirecutter. Finally, the Complaint cites two examples of the Times’s attempts to probe ChatGPT about “Wirecutter,” a section of the Times that recommends products in exchange for a “commission” from manufacturers. Compl. ¶ 128. In both, the Times asked ChatGPT about a specific Wirecutter recommendation, *see id.* ¶ 134, and ChatGPT responded by directing the user to Wirecutter itself and providing a short, non-verbatim summary of the recommendation. *Id.* ¶ 130 (including hyperlink); *id.* ¶ 134 (urging user to “check [Wirecutter’s] latest reviews”).

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).

IV. ARGUMENT

This Motion seeks dismissal of four claims. **(1)** OpenAI seeks partial dismissal of Count I (Direct Copyright Infringement) to the extent it is based on acts of reproduction 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 failure to identify the CMI at issue, failure to allege OpenAI “remove[d]” CMI from any datasets or outputs, failure to allege “distribution,” and failure to plead facts that suggest that OpenAI acted with scienter. *Infra* Section IV(C). And **(4)** OpenAI seeks full dismissal of Count VI (Unfair

article in question is available for free. *See id.* ¶ 118; compare Hurubie Meko, *The Precarious, Terrifying Hours After a Woman Was Shoved Into a Train*, N.Y. Times (May 25, 2023), <https://www.nytimes.com/2023/05/25/nyregion/subway-attack-woman-shoved-manhattan.html?smid=url-share>, with Hurubie Meko, *The precarious, terrifying hours after a woman was shoved into a train*, Seattle Times (May 27, 2023), <https://www.seattletimes.com/nation-world/nation/the-precarious-terrifying-hours-after-a-woman-was-shoved-into-a-train/>.

Competition by Misappropriation) on grounds of Copyright Act preemption. *Infra* Section IV(D).

A. The Times 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. ¶¶ 160, 162. That claim appears to hinge on allegations regarding **(1)** construction of the “WebText” database and OpenAI’s use of that dataset to train GPT-2, *see id.* ¶ 85; **(2)** construction of an “expanded version of the WebText dataset” called “WebText2,” *see id.* ¶ 87; and **(3)** use of WebText2 and Common Crawl to train GPT-3, *see id.* ¶¶ 87–88. Because all those activities occurred more than three years ago, *supra* 7, any claims based on them are time-barred, 17 U.S.C. § 507(b) (three-year limitations period).³³ Those 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. ¶ 179. This claim relies on the doctrine of “contributory infringement,” a species of secondary liability that the Supreme Court has defined by reference to the Patent Act. 35 U.S.C. § 271(b); *Sony*, 464 U.S. at 435–41. 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 the Complaint,

³³ These claims are time-barred regardless of whether the discovery rule applies, as the Times discovered or with reasonable diligence should have discovered these activities prior to December 27, 2020.

discussed above. *See supra* 12. To proceed with Count IV, the Times must allege that OpenAI “had knowledge” of the Times’s creation of those outputs. *Dow Jones*, 2023 WL 2561588, at *3.

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.” *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 934 (2005). As the Supreme Court explained in *Commil USA, LLC v. Cisco Sys., Inc.*, contributory patent infringement requires more than allegations that a defendant knew there “might” be infringement. 575 U.S. 632, 642 (2015). Instead, a plaintiff must allege that the defendant had *actual knowledge* of *specific* infringements.

Courts have consistently applied this same approach to copyright claims. Pleading a contributory copyright claim requires 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 (9th Cir. 2013).³⁴ As the Fourth Circuit recently explained, this follows directly from *Commil*, as well as other foundational Supreme Court cases like *Sony*, 464 U.S. 417, and *Grokster*, 545 U.S. 913. *See BMG Rights Mgmt. (US) LLC v. Cox Commc’ns, Inc.*, 881 F.3d 293, 308–10 (4th Cir. 2018) (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.³⁵

Here, the only allegation supporting the Times’s contributory claim states that OpenAI “had reason to know of the direct infringement by end-users” because of its role in “developing,

³⁴ 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”).

³⁵ *Lefkowitz v. John Wiley & Sons*, No. 13-cv-6414, 2014 WL 2619815, at *11 (S.D.N.Y. June 2, 2014) (dismissing claim); *State Street Global Advisors Trust Co. v. Visbal*, 431 F. Supp. 3d 322, 358 (S.D.N.Y. 2020) (same); *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).

testing, and troubleshooting” its products. Compl. ¶ 180. But “generalized knowledge” of “the possibility of infringement” is not enough. *Luvdarts*, 710 F.3d at 1072. The Complaint does not allege OpenAI “investigated or would have had reason to investigate” the use of its platform to create copies of Times articles. *Popcornflix.com*, 2023 WL 571522, at *6. Nor does it suggest that OpenAI had any reason to suspect this was happening. Indeed, OpenAI’s terms expressly prohibit such uses of its services. *Supra* note 8. And even if OpenAI *had* investigated, nothing in the Complaint explains how it might evaluate whether these outputs were acts of copyright infringement or whether their creation was authorized by the copyright holder (as they were here).

C. The DMCA Claim Fails for Multiple Independent Reasons

Count V is a claim for violation of Section 1202(b) of the Copyright Act, which 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 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 removal) so that, if their works *do* proliferate on the internet, the public will be able to trace those works back to their owner. 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). 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 Times Did Not Specify the CMI at Issue*

Count V should be dismissed at the outset for failure to specify the CMI at issue. The

Complaint’s relevant paragraph fails to state what CMI is included in what work, and simply repeats the statutory text. Compl. ¶ 182 (alleging “one or more forms of [CMI]” and parroting language of Section 1202(c)).³⁶ The only firm allegation states that the Times placed “copyright notices” and “terms of service” links on “every page of its websites.” Compl. ¶ 125. But, at least for some articles, it did not.³⁷ And when it did, the information was not “conveyed in connection with” the works, 17 U.S.C. § 1202(c) (defining CMI), but hidden in small text at the bottom of the page.³⁸ Judge Orrick of the Northern District of California rejected similar allegations as deficient in another recent AI case. *Andersen v. Stability AI Ltd.*, No. 23-cv-00201, 2023 WL 7132064, at *11 (N.D. Cal. Oct. 30, 2023) (must plead “exact type of CMI included in [each] work”).³⁹

2. *The Training-Based Section 1202 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. ¶ 184. 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).

The Complaint also fails to plausibly allege that any CMI was removed. The Times advances three theories of removal: **(1)** removal of CMI when OpenAI allegedly “scraped” articles from the Times’s website; **(2)** removal of CMI “from third-party datasets,” *i.e.*, Common Crawl; and **(3)** removal of CMI during “the training process,” which the Times alleges “does not preserve

³⁶ *Design Pics Inc. v. PBH Network, Inc.*, No. 20-cv-1096, 2020 WL 8413512, at *4 (E.D.N.Y. Oct. 27, 2020) (dismissing CMI claim based on “conclusory, boilerplate parroting of the statutory text”).

³⁷ See, e.g., John Branch, *Snow Fall: The Avalanche at Tunnel Creek*, N.Y. Times, <https://www.nytimes.com/projects/2012/snow-fall/index.html#/?part=tunnel-creek> (last accessed Feb. 11, 2024); see also Compl. ¶¶ 104–05 & n.28 (citing this article).

³⁸ See, e.g., Pete Wells, *As Not Seen on TV*, N.Y. Times (Nov. 13, 2012), <https://www.nytimes.com/2012/11/14/dining/reviews/restaurant-review-guys-american-kitchen-bar-in-times-square.html>; see also Compl. ¶¶ 106–07 & n.29 (citing and quoting this article).

³⁹ See also *Wood v. Observer Holdings, LLC*, No. 20-cv-07878, 2021 WL 2874100, at *8 (S.D.N.Y., July 8, 2021) (terms on “separate website” not CMI); *GC2 Incorporated v. Int’l Game Tech. PLC*, 255 F. Supp. 3d 812, 821–22 (N.D. Ill. 2017) (“terms of use notice near a copyrighted work” not “conveyed in connection with” work).

any [CMI]” “by design.” Compl. ¶¶ 184, 187. As a preliminary matter, each theory fails for failure to allege *what* CMI was removed or “not preserve[d],” which is particularly damning as the Times concedes that some CMI *was* preserved.⁴⁰ Judge Martínez-Olguín of the Northern District of California recently held an identical set of allegations to be insufficient. *See Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03416, 2024 WL 557720, at *4 (N.D. Cal. Feb. 12, 2024) (allegation that “training process does not preserve any CMI” “by design” was “conclusory”).

Each theory fails separately as well. The first theory fails because, while the Times states that OpenAI “scraped” its articles “directly from [its] websites,” none of the specific allegations actually suggest OpenAI designed its alleged “scrap[ing]” process to omit CMI. Compl. ¶ 184.⁴¹ And the only allegations about OpenAI “scraping” articles from Times websites relate to the creation of WebText, which occurred over three years before this lawsuit. *Supra* Section IV.A.⁴² The second theory fails because the Complaint lacks allegations about the inclusion (or exclusion) of the Times’s CMI in any “third-party datasets” like Common Crawl,⁴³ much less about OpenAI scrubbing any 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 the third theory fails because there is no allegation in the Complaint supporting the conclusion that the “training process” excludes CMI “[b]y design.” Compl. ¶ 187; *see also Tremblay*, 2024 WL 557720, at *4.

Moreover, the Times fails to allege facts that could show how the alleged CMI removal could “induce, enable, facilitate, or conceal an infringement” of copyright—much less how

⁴⁰ *See, e.g.*, Compl. ¶ 106 (ChatGPT responding to query naming “Pete Wells” with a completion correctly identifying the publication date of Wells’ article); *see also id.* ¶ 104 (same, for title).

⁴¹ *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999), *rev’d on other grounds by* 336 F.3d 811 (9th Cir. 2003) (no Section 1202 liability vs. search engine “crawler [that] did not include [CMI]”).

⁴² The Times suggests that OpenAI’s “Browse with Bing” feature “scrap[es] Times Works from The Times’s websites,” Compl. ¶ 185, but the Complaint does not include a single allegation supporting that conclusion, *see supra* 13 & n.32 (noting that Browse with Bing fetched content from third-party sites, not the Times’s website).

⁴³ 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).

OpenAI could have “reasonable grounds to know” it would. 17 U.S.C. § 1202(b). The “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. ¶ 59. As Judge Martínez-Olguín explained, it is far from obvious how “the alleged removal of CMI in an internal database [could] enable infringement.” *Tremblay*, 2024 WL 557720, at *4 (dismissing claim).⁴⁴

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 the Times’s CMI in model outputs, Compl. ¶¶ 185–86; and **(2)** by displaying those outputs via ChatGPT or its API, OpenAI violated Section 1202(b)(3)’s prohibition on “distribut[ing]” works “knowing that [CMI] has been removed,” Compl. ¶ 189. Neither theory states a claim for relief.

As a preliminary matter, the Times’s 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).⁴⁵ But “mere public display” of outputs is all the Complaint alleges. *See, e.g.*, Compl. ¶ 102.

Regardless, this “output” theory fails because the outputs alleged in the Complaint are not wholesale copies of entire Times articles. They are, at best, reproductions of *excerpts* of those articles, some of which are little more than collections of scattered sentences. *Supra* 12. If the absence of CMI from such excerpts constituted a “removal” of that CMI, then DMCA liability

⁴⁴ *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.”).

⁴⁵ *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.”).

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. *See supra* note 22 (example of the Times including 200-word block quote in book review).

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]”).⁴⁶ As the Times has not alleged that OpenAI reproduced entire articles, the output-based claim fails.

Even setting that aside, the Times’s 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 2–126. As shown in the exhibit, the “Actual text from NYTimes” contains no information that could qualify as CMI. *See, e.g., id.* at 3; 17 U.S.C. § 1202(c) (defining CMI). So too for the ChatGPT outputs featured in the Complaint, which request the “first [and subsequent] paragraph[s]” from Times articles. *See, e.g., Compl.* ¶¶ 104, 106, 118, 121. None of those “paragraphs” contains any CMI that OpenAI could have “removed.”

4. *The Times Fails to Allege a CMI-Based Injury*

Count V separately fails for lack of standing. “[T]o have standing” to sue for a DMCA violation, the Times “must show that [it] was injured by that violation.” *Steele v. Bongiovi*, 784

⁴⁶ *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”).

F. Supp. 2d 94, 97–98 (D. Mass. 2011); *see also* 17 U.S.C. § 1203(a). Here, the Complaint’s “Harm to the Times” section relates entirely to its inability to receive speculative licensing revenue, *see* Compl. ¶¶ 155–56, and the possibility that ChatGPT will “divert readers,” *see id.* ¶ 157. Neither injury has any nexus to CMI. Nor is there any imaginable harm here: because all of the Complaint’s outputs were either generated using the original Times article itself, *see* Ex. J, or referenced the Times by name, *see, e.g.*, Compl. ¶ 104, any user who encountered those outputs would have no doubt as to the provenance of the text and could easily find it on the Times’s website (as ChatGPT often invites them to do, *see id.* ¶¶ 106, 134). *Cf. Kelly*, 77 F. Supp. 2d at 1122 (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”).

D. The “Misappropriation” Claim Is Preempted by the Copyright Act

Count VI, for “misappropriation” under New York law, appears to raise two distinct theories. The first suggests that OpenAI engages in “unfair competition” by using “Times content to train models that produce informative text of the same general type and kind that [t]he Times produces.” Compl. ¶ 195 (the “Text Claim”). The second suggests that OpenAI harms the Times because ChatGPT can respond to user queries about Wirecutter recommendations. *Id.* ¶ 194 (the “Recommendations Claim”). Both are preempted by the Copyright Act. 17 U.S.C. § 301(a).

1. *The Text Claim Fails*

The Text Claim is preempted by Section 301 of the Copyright Act, which “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) Times 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. One aspect of the Text Claim is based on the use of content to “train models,” Compl. ¶ 195, which is one of the stated bases of the copyright claim, *id.* ¶¶ 64, 161–62. The other is based on outputs that are “the same or similar to content published by [t]he Times,” *id.* ¶ 194–95, which is the other stated basis of the copyright claim, *see id.* ¶ 163. Because the Text Claim is based on the same allegations as the copyright claims, the claim is preempted.⁴⁷

2. *The Recommendations Claim Fails*

The Recommendations Claim focuses on Wirecutter, a Times-owned website that publishes articles to guide readers through difficult purchase decisions. *See, e.g.*, Compl. ¶¶ 130, 134. Wirecutter makes “the vast majority of its revenue” through “commission[s],” which it earns when users click “affiliate links” in its articles and purchase a product. *Id.* ¶¶ 128–29. According to the Times, by complying with requests to identify Wirecutter-recommended products, ChatGPT reduces the need for users to “visit the [] Wirecutter article and click on [these] links,” *id.*, thus “depriving [t]he Times of the opportunity to receive referral revenue,” *id.* ¶ 194.

This claim is almost identical to the “misappropriation” claim the Second Circuit rejected in *Barclays*. 650 F.3d 876 (2d Cir. 2011). That case concerned stock recommendations which banks provided to paying clients as a mechanism for generating commissions. *Id.* at 880–82. The

⁴⁷ *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).

defendant “compile[d]” those recommendations and provided them “to their own subscribers for a fee.” *Id.* at 882. The banks brought a “misappropriation” claim, stressing that they “spend hundreds of millions of dollars annually” generating the recommendations, and that the defendant “seriously threatens their ability to justify [that] expense.” *Id.* at 881, 885–86.

The Second Circuit first held that the claim satisfied the conditions for Copyright Act preemption: **(1)** the “facts of the Recommendations,” while unprotectable, fell within the subject matter of copyright; and **(2)** the banks claimed the right to control the reproduction of those recommendations, which falls within the “general scope” of copyright. *Id.* at 902. The Court, however, explained that its precedents recognized a “narrow” preemption exception for “hot news” claims endorsed by the Supreme Court in the *Int’l News Serv. v. Associated Press (INS)*, 248 U.S. 215 (1918) case. *Barclays*, 650 F.3d at 896–98.⁴⁸ Accordingly, the merits of the claim depended on whether it fell into the “narrow” category of “INS-type non-preempted claims.” *Id.* at 902–03.

The Circuit held it did not. First, while “INS-type” claims involve defendants who “sell [purloined] news *as though the defendant itself had gathered it*,” the *Barclays* defendant was “selling the information with *specific attribution* to the issuing [bank].” *Id.* (emphasis added). For that reason, there was no “meaningful difference” between the defendant’s republication of the recommendations and a “member[] of the traditional news media” reporting on information “with proper attribution.” *Id.* at 903–04. Second, INS-type claims concern defendants who appropriate material (*e.g.*, news) that the defendants “*acquire* through efforts akin to reporting.” *Id.* at 903. But the banks’ claim sought “only to protect their Recommendations, something they *create* using their expertise and experience”—which brought the claim closer to copyright’s exclusive domain. *Id.* For those reasons, the defendant’s “service—which collects, summarizes, and disseminates

⁴⁸ *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*).

the news of the [banks'] Recommendations—[was] not the kind of 'INS-like' product that could support a non-preempted cause of action for misappropriation." *Id.* at 905.

The Recommendations Claim fails for the same reasons. First, it satisfies the preemption conditions: it **(1)** relates to facts within copyright's "subject matter;" and **(2)** asserts the right to control reproduction of those facts, which falls within copyright's "general scope." *Id.* at 902.

Second, to the extent the Times has attempted to plead an "INS-type non-preempted claim[]" for misappropriation, *id.* at 902, it has failed to do so for the reasons discussed in *Barclays*. OpenAI is not "selling the Recommendation[s] 'as its own.'" *Id.* at 903. It provides "specific attribution to" Wirecutter, which means this is not an *INS* claim. *Id.* Moreover, Wirecutter recommendations are not facts that the Times "*acquire[s]* through efforts akin to reporting;" rather, the Times "seek[s] only to protect [its] Recommendations, [which it] *create[s]* using [its] expertise and experience." *Id.* The claim is therefore governed exclusively by copyright law, under which the quantum of the Times's "investment" is irrelevant to its right to monopolize the results. Compl. ¶ 196; *Moody's*, 808 F.2d at 207 ("[T]o grant copyright protection based merely on the 'sweat of the author's brow' would risk putting large areas of factual research material off limits and threaten the public's unrestrained access to information."); *Barclays*, 650 F.3d at 886 (rejecting claim despite allegation that defendant "seriously threatens [banks'] ability to justify the expense of maintaining [] extensive research operations").

V. CONCLUSION

For the foregoing reasons, the Court should **(1)** dismiss Counts IV, V, and VI; **(2)** dismiss Count I to the extent it is based on acts outside the limitations period; and, **(3)** if Count V survives, dismiss that count to the extent it is based on acts outside the limitations period.

Dated: February 26, 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
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Allison L. Stillman
alli.stillman@lw.com
Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.751.4864

By: /s/ Joseph C. Gratz

MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)*
jgratz@mofocom
Vera Ranieri (*pro hac vice forthcoming*)
vraniieri@mofocom
425 Market Street
San Francisco, CA 94105
Telephone: 415.258.7522

Allyson R. Bennett (*pro hac vice forthcoming*)
abennett@mofocom
Rose S. Lee (*pro hac vice*)
roselee@mofocom
707 Wilshire Boulevard, Suite 6000
Los Angeles, CA 90017-3543
Telephone: 213.892.5454

Attorneys for OpenAI Defendants

* The parties use electronic signatures with consent in accordance with Rule 8.5(b) of the Court's ECF Rules.