**BEFORE THE UNITED STATES JUDICIAL PANEL ON
MULTIDISTRICT LITIGATION**

IN RE OPENAI COPYRIGHT LITIGATION	MDL No.

**BRIEF IN SUPPORT OF OPENAI'S MOTION FOR TRANSFER OF ACTIONS
PURSUANT TO 28 U.S.C. § 1407 FOR COORDINATED OR CONSOLIDATED
PRETRIAL PROCEEDINGS**

2832519

TABLE OF CONTENTS

2832519

TABLE OF AUTHORITIES

Page(s)

Federal Cases

2832519

I.   INTRODUCTION

Between June 2023 and August 2024, twelve lawsuits were filed in the United States against the artificial intelligence company OpenAI[1] asserting claims for direct copyright infringement, 17 U.S.C. § 501, and/or violations of the Digital Millennium Copyright Act, 17 U.S.C. § 1202, (collectively referred to herein as the "OpenAI Copyright Litigation").[2] *See* Schedule of Actions, filed herewith ("Schedule A"). Each of these lawsuits arises from the same underlying allegations: that OpenAI used plaintiffs' copyrighted works to train certain large language models ("LLMs") which, in turn, produced outputs that summarize or quote from those works through the popular LLM-powered tool ChatGPT.[3] Among other defenses, OpenAI asserts that its activity is protected by the doctrine of fair use.[4]

There has been a limited consolidation of certain lawsuits, but a minimum of eight actions remain pending and are unlikely to be further consolidated. These eight actions are spread across five judges in two districts and involve 11 sets of plaintiffs' counsel. OpenAI now moves for centralization of these actions under 28 U.S.C. § 1407 to avoid duplicative discovery—including especially multiple redundant depositions of OpenAI's witnesses—and

---

[1] Most lawsuits name various OpenAI entities as defendants, including: OpenAI Inc., OpenAI OpCo LLC, OpenAI GP LLC, OpenAI LLC, OpenAI Global LLC, OAI Corporation LLC, OpenAI Holdings LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC. *See, e.g.*, *Authors Guild et al. v. OpenAI, Inc. et al.*, Case No. 1:23-cv-08292-SHS, Dkt. 69 (S.D.N.Y. Feb. 5, 2024) ("*Authors Guild*"). As OpenAI has maintained in the underlying actions, the sole entity relevant to each lawsuit is OpenAI OpCo LLC. As used in this brief, however, the term "OpenAI" refers to all OpenAI entities named as defendants.

[2] There have also been five additional and substantially similar lawsuits filed internationally: three in Canada, one in India, and one in Germany. These lawsuits, most of which were filed just weeks before this motion, are not presently part of this motion or included in Schedule A.

[3] All plaintiffs allege that ChatGPT improperly quotes or summarizes from copyrighted works, but some plaintiffs' copyright claims relate only to OpenAI's alleged use of copyrighted works to train its LLM, and not outputs.

[4] *See, e.g., id.* Dkt. 75; *see also* OpenAI, *OpenAI and journalism* (January 8, 2024) https://openai.com/index/openai-and-journalism/. In addition to its fair use defense, OpenAI denies many of the allegations made in each complaint, including that ChatGPT outputs are infringing, and has raised numerous other defenses to the various plaintiffs' claims.

2832519

inconsistent pretrial rulings on important and emerging questions of law that will inform the business and research practices of OpenAI and other generative AI companies for years to come.

Earlier in this litigation—when there were only three cases to coordinate and two were before the same judge—OpenAI was optimistic that it could coordinate the cases without transfer or Section 1407 centralization. As of this filing, however, five more cases have been filed against OpenAI, and there is no reasonable prospect for coordination or streamlining the litigation through other means. OpenAI's extensive efforts to negotiate a deposition coordination protocol among just a subset of the 11 groups of plaintiffs' counsel have not been successful. And OpenAI's request for entry of a deposition coordination protocol has been denied as of the date of this filing. Magistrate judges in the Northern District of California and the Southern District of New York have also issued inconsistent guidance on discovery, and two district court judges within the Southern District of New York recently issued conflicting opinions on OpenAI's motions to dismiss identical DMCA claims in separate—but near-identical—cases. Every traditional consideration under Section 1407 counsels in favor of centralization.

Here, centralization makes most sense in the Northern District of California. OpenAI's headquarters and witnesses are located within the Northern District of California. The individuals responsible for training OpenAI's LLMs are located in the Northern District of California, and all source code and training data inspections are taking place in the Northern District of California. Accordingly, OpenAI respectfully requests that this Panel order transfer of all actions listed in Schedule A to the Northern District of California for pretrial consolidation.

II.   STATEMENT OF FACTS

A.   California Consolidated Class Action

On June 28, 2023, authors Paul Tremblay and Mona Awad, represented by the Saveri Law Firm and attorney Matthew Butterick, filed a putative class action against OpenAI in the Northern District of California, asserting claims for copyright infringement under 17 U.S.C. § 501, DMCA violations under 17 U.S.C. § 1202(b), and other related California state law claims arising from OpenAI's alleged use of their works to train its LLMs. *Tremblay et al. v. OpenAI et*

2832519

al., Case No. 3:23-cv-03223, Dkt. 1 (N.D. Cal. June 28, 2023). The class was defined to include all United States copyright holders whose works were used to train OpenAI's LLMs. *Id.* Approximately one week later, on July 7, 2023, the same attorneys filed a second putative class action against OpenAI asserting identical claims on behalf of an identical class. *Silverman, et al.* v. *OpenAI, Inc., et al.*, Case No. 3:23-cv-03416, Dkt. 1 (N.D. Cal. July 7, 2023). On September 8, 2023, a third set of authors represented by Cafferty Clobes Meriwether & Sprengel LLP and Ventura Hersey & Muller, LLP, filed yet another near-identical putative class action against OpenAI. *Chabon et al.* v. *OpenAI, Inc. et al.*, Case No. 3:23-cv-04625, Dkt. 1 (N.D. Cal. Sept. 8, 2023). Like the two prior putative class actions, the third class action asserted claims for copyright infringement and related California state law claims arising from OpenAI's alleged use of their works to train its LLMs on behalf of an identically defined class. *Id.*

Each of these cases was assigned to Judge Araceli Martinez-Olguin who, on November 9, 2023, ordered that the two later-filed cases be consolidated with the *Tremblay* action (the "California Consolidated Class Action"). *In re OpenAI ChatGPT Litigation*, Case No. 3:23-cv-03223, Dkt. 74 (N.D. Cal. Nov. 9, 2023) ("*In re ChatGPT*"). On March 13, 2024, the California Consolidated Class Action Plaintiffs filed a First Amended Consolidated Complaint asserting claims for direct copyright infringement and California unfair competition,[5] again on behalf of all United States copyright holders whose works had been used to train OpenAI's LLMs. *Id.* Dkt. 120.

As of the date of this filing, the close of fact discovery in the California Consolidated Class Action is set for April 28, 2025. Declaration of Michelle Ybarra in Support of OpenAI's Motion for Transfer of Actions Pursuant to 28 U.S.C. § 1407 for Coordinated or Consolidated Pretrial Proceedings ("Ybarra Decl.") Ex. 2 (Nov. 26, 2024 Hr'g. Tr., *In re ChatGPT*) at 18:15-16. No depositions have been taken, and document discovery remains ongoing.

---

[5] Judge Martinez-Olguin granted OpenAI's motion to dismiss the latter claim on July 30, 2024. *In re ChatGPT,* Dkt. 162.

2832519

B.   **New York Consolidated Class Action**

On September 19, 2023, the Authors Guild and several authors of fictional works, represented by Lieff Cabraser Heimann & Bernstein LLP ("Lieff Cabraser") and Cowan Debaets Abrahams & Sheppard LLP ("Cowan Debaets"), filed a putative class action against OpenAI in the Southern District of New York. *Authors Guild*, Dkt. 1. The complaint asserted claims for direct, vicarious, and contributory copyright infringement arising from OpenAI's alleged use of their works to train its LLMs. *Id.* The *Authors Guild* class was defined to include all United States "sole authors" of "[a]ny work of fiction that has sold at least 5,000 copies" that was used to train OpenAI's LLMs. *Id.* ¶¶ 311-312. On December 5, 2023, the *Authors Guild* Plaintiffs amended their complaint to add claims against Microsoft Corporation ("Microsoft"). *Id.* Dkt. 40.

On November 11, 2023, the non-fiction writer Julian Sancton (and, upon amendment, other non-fiction writers including Jonathan Alter), represented by Susman Godfrey L.L.P. ("Susman Godfrey"), filed a putative class action asserting claims for direct and contributory copyright infringement against OpenAI and Microsoft. *Alter et al. v. OpenAI et al.*, Case No. 1:23-cv-10211, Dkts. 1, 26 (S.D.N.Y. Nov. 21, 2023 and Dec. 19, 2023) ("*Alter*"). The *Alter* class was defined to include all owners of copyrighted and registered non-fiction literary works that were used to train OpenAI's LLMs. *Id.* Dkt. 26. On January 5, 2024, writers Nicholas Basbanes and Nicholas Ngagoyeanes, represented by Grant Herrmann Schwartz & Klinger LLP, filed another near-identical putative class action against OpenAI and Microsoft on behalf of themselves and similarly situated authors. *Basbanes, et al. v. Microsoft, et al.*, Case No. 1:24-cv-00084, Dkt. 1 (S.D.N.Y. Jan. 5, 2024) ("*Basbanes*").

Each of these cases was assigned to Judge Sidney Stein who, on January 22, 2024, consolidated *Alter* into *Authors Guild* (the "New York Consolidated Class Action") pursuant to the parties' stipulation. *Authors Guild*, Dkt. 56. That stipulation included a compromise from OpenAI that it would not seek to transfer the New York Consolidated Class Action or *The New York Times* action described below—at that time, the only two cases in that district (besides *Basbanes*, which was soon consolidated), both of which were assigned to Judge Stein—away

4

2832519

from the Southern District of New York under Section 1404. *Id.* at 3. The stipulation included an express carve out for a transfer motion under Section 1407 in the event that more similar cases were filed. *Id.*

On February 5, 2024, the *Authors Guild* and *Alter* Plaintiffs filed a First Consolidated Class Action Complaint asserting claims for direct, vicarious, and contributory copyright infringement against OpenAI and Microsoft on behalf of separate classes of fiction and non-fiction authors. *Id.* Dkt. 69. On February 6, 2024, Judge Stein consolidated *Basbanes* into the New York Consolidated Class Action. *Id.* Dkt. 70. On September 26, 2024, the *Basbanes* Plaintiffs stipulated to dismiss their class claims, proceed individually, and stay their action pending a decision on class certification in the consolidated action. *Basbanes*, Dkt. 103.

On February 8, 2024, the California Consolidated Class Action Plaintiffs filed a motion in the Northern District of California to "Enjoin Defendants and Their Counsel from Proceeding in Substantially Similar Cases in the Southern District of New York," *In re ChatGPT*, Dkt. 98, which Judge Martinez-Olguin denied on March 1, 2024, *id.* Dkt. 118. On February 12, 2024, the California Consolidated Class Action Plaintiffs filed a motion to intervene and dismiss or, in the alternative, stay and transfer, the New York Consolidated Class Action under the first-to-file rule. *Authors Guild*, Dkt. 71. Judge Stein denied that motion on April 1, 2024. *Id.* Dkt. 100. The California Consolidated Class Action Plaintiffs filed a Notice of Interlocutory Appeal to the United States Court of Appeals for the Second Circuit, *id.* Dkt. 104, and then voluntarily dismissed that appeal several months later, *id.* Dkt. 215.

On August 6, 2024, Judge Stein referred the New York Consolidated Class Action to Magistrate Judge Ona Wang for general pretrial purposes, including "scheduling, discovery, non-dispositive pretrial motions, and settlement[.]" *Id*. Dkt. 177. As of the date of this filing, the close of fact discovery is set for April 30, 2025, though Judge Wang has indicated that deadline may need to be extended. No depositions have been taken, and written discovery remains ongoing.

2832519

C.   **New York Consolidated News Cases**

On December 27, 2023, the New York Times Company, represented by Susman Godfrey[6] and Rothwell, Figg, Ernst & Manbeck, P.C. ("Rothwell Figg"), filed a lawsuit against OpenAI and Microsoft in the Southern District of New York asserting claims for copyright infringement, DMCA violations, trademark dilution, and common law misappropriation. *The New York Times v. Microsoft et al.*, Case No, 1:23-cv-11195, Dkt. 1 (S.D.N.Y. Dec. 27, 2023) ("*NY Times*"). Each of these claims arose from OpenAI's alleged training of its LLMs on approximately three million asserted New York Times works. *Id.*

On February 23, 2024, the California Consolidated Class Action Plaintiffs moved to intervene and dismiss or, in the alternative, stay and transfer *The New York Times* in a motion substantially identical to the one filed by the same Plaintiffs in the New York Consolidated Class action. *Id.* Dkt. 47. Judge Stein denied the California Plaintiffs' motion in the same order in which he denied their motion in the New York Consolidated Class Action. *Id.* Dkt. 84. On May 20, 2024, The New York Times moved to amend its complaint to add approximately seven million more asserted works to those in the original complaint, thereby resulting in an operative complaint asserting approximately 10 million works. *Id.* Dkts. 118, 170.

On April 30, 2024, the Daily News, LP and seven other regional newspapers, also represented by Rothwell Figg (but not Susman Godfrey), filed a lawsuit against OpenAI and Microsoft in the Southern District of New York. *Daily News et al. v. Microsoft et al.*, Case No. 1:24-cv-03285, Dkt. 1 (S.D.N.Y. April 30, 2024) ("*Daily News*"). The complaint asserts claims almost identical to those brought by The New York Times and alleges that OpenAI used approximately 8 million asserted works to train its LLMs.

On June 27, 2024, the Center for Investigative Reporting ("CIR"), an investigative journalism organization represented by the law firm Loevy & Loevy, filed a lawsuit against OpenAI and Microsoft in the Southern District of New York. *Center for Investigative Reporting*

---

[6] The groups of lawyers from Susman Godfrey representing The New York Times Company and the class plaintiffs in the *Authors Guild* action are distinct.

v. OpenAI et al., Case No. 1:24-cv-04872, Dkt. 1 (S.D.N.Y. June 27, 2024) ("*CIR*"). The complaint asserts similar claims to those brought by The New York Times and Daily News Plaintiffs, for direct copyright infringement, contributory copyright infringement, and violation of Section 1202 of the DMCA. *The New York Times*, *Daily News*, and *CIR* were all assigned to Judge Stein and related to one another.

On June 13, 2024—two weeks before CIR filed its lawsuit—OpenAI moved to consolidate the *Daily News* and *New York Times* cases and extend the discovery cut-off for both cases by approximately six months in light of the massive volume of asserted works between the two near-identical lawsuits. *Daily News*, Dkt. 87. The *Daily News* and *New York Times* Plaintiffs responded with a proposal calling for some degree of coordinated depositions, cross-production of Defendants' documents and a coordinated discovery schedule extending fact discovery by only two months beyond the previously set deadline in the *New York Times* case. *Id.* Dkt. 102.

On August 6, 2024, Judge Stein referred the *New York Times*, *Daily News*, and *CIR* actions (together, the "News Cases") to Magistrate Judge Ona Wang for general pretrial purposes, including "scheduling, discovery, non-dispositive pretrial motions, and settlement[.]" *NY Times*, Dkt. 162. On September 12, 2024, Judge Wang ordered the consolidation of the *New York Times* and *Daily News* lawsuits. Ybarra Decl. Ex. 3 (Sept. 12, 2024 Hr'g. Tr., *Authors Guild*) at 72:4-8. On October 4, 2024, OpenAI moved to consolidate the *CIR* lawsuit with the other consolidated news cases. *CIR*, Dkt. 93. Judge Wang granted that motion on October 31, 2024 and placed all three cases on the same discovery schedule (the "Consolidated News Actions"). *NY Times*, Dkt. 304. In consolidating these cases, however, Judge Wang did not order that a consolidated complaint be filed, nor specify the form that consolidation was to take, nor degree of intended coordination. *Id.* As of the date of this filing, discovery for each of the three cases has largely proceeded independently—the plaintiffs have served distinct, but heavily overlapping, discovery requests and have met-and-conferred separately about OpenAI's responses. The *New York Times* and *Daily News* have coordinated their source code and training data inspection efforts, but *CIR* has not joined them.

2832519

As of the date of this filing, the close of fact discovery is set for April 30, 2025, though Judge Wang has indicated that deadline may need to be extended. No depositions have been taken, and written discovery remains ongoing.

D.   New York DMCA Cases

On February 28, 2024, online news outlets Raw Story Media, Inc. and Alternet Media Inc, represented by Loevy & Loevy, filed a lawsuit against OpenAI in the Southern District of New York asserting a single claim for violation of the DMCA, 17 U.S.C. § 1202(b), arising from OpenAI's alleged use of their works to train its LLMs and purported removal of copyright management information from those works in the course of that training. *Raw Story et al. v. OpenAI*, Case No. 1:24-cv-01514, Dkt. 1 (S.D.N.Y. Feb. 28, 2024) ("*Raw Story*"). On the same day, the same attorneys filed a second DMCA lawsuit against both OpenAI and Microsoft in the Southern District of New York, this time on behalf of another online news organization, The Intercept Media, Inc. *The Intercept Media, Inc. v. OpenAI et al.*, Case No. 1:24-cv-01515, Dkt. 1 (S.D.N.Y. Feb. 28, 2024) ("*Intercept*"). The *Intercept* complaint differed from the *Raw Story* complaint only in that it added Microsoft as a defendant and asserted separate claims against OpenAI for alleged violations of 17 U.S.C. § 1202(b)(1) and (3).

Both the *Raw* Story and *The Intercept* Plaintiffs attempted to relate their cases to the then-recently filed *New York Times* action before Judge Stein. But Judge Stein declined relation, despite the near-identical nature of the three cases' DMCA claims. *Raw Story* was assigned to Judge Colleen McMahon and *The Intercept* case was assigned to Judge Jed Rakoff. On March 19, 2024, OpenAI proposed that Judge Rakoff relate *The Intercept* to *Raw Story* in light of the material identicality of the two complaints. On March 20, 2024, Judge Rakoff informed the parties that he had determined that the cases were not related. Ybarra Decl. Ex. 4.

On November 11, 2024, Judge McMahon granted OpenAI's motion to dismiss the *Raw Story* complaint on grounds that the plaintiffs had not adequately alleged injury in fact and thus lacked Article III standing. *Raw Story*, Dkt. 117 (instructing plaintiffs to seek leave to file any amendment). The *Raw Story* Plaintiffs moved for leave to amend on November 21, 2024, *id.*

Dkt. 119, which OpenAI will oppose on December 20, 2024, *id.* Dkt. 120. On November 22, 2024, Judge Rakoff granted-in-part and denied-in-part OpenAI's very similar motion to dismiss the *Intercept* complaint, permitting the Plaintiffs' DMCA claim under 17 U.S.C. § 1202(b)(1) to survive and rejecting OpenAI's arguments about insufficiently pled injury.

Discovery in *Raw Story* was ongoing before the dismissal, but as of the date of this filing discovery is not active in either of these cases because there is no effective complaint in *Raw Story* and discovery remains stayed in *Intercept*.

E.   California YouTube Class Action

On August 2, 2024, Plaintiff David Millette, represented by the law firm Bursor & Fisher P.A., filed a lawsuit against OpenAI in the Northern District of California, asserting California state law claims on behalf of himself and a putative class of YouTube content creators whose works were used to train OpenAI LLMs. *Millette et al. v. OpenAI*, Case No. 3:24-cv-04710, Dkt. 1 (N.D. Cal. Aug. 2, 2024) ("*Millette*"). On October 18, 2024, Millette filed an amended complaint adding a named plaintiff and claims for direct copyright infringement under federal law and unfair competition under Massachusetts law, Mass Gen. Law Ch. 93A *et seq. Id.* Dkt. 47. The case is currently before Judge Edward Davila. The parties have entered into a stipulation pursuant to which OpenAI will move to dismiss the amended complaint on December 16, 2024 and discovery is to remain stayed pending resolution of that motion. *Id.* Dkt. 53.

F.   Discovery Status and Coordination Efforts

Document and written discovery in this case has already been extraordinarily voluminous and duplicative. The California Class Action Plaintiffs have served a total of 80 document requests, 16 Interrogatories, and 178 Requests for Admission ("RFAs") on OpenAI. Ybarra Decl. ¶3. The New York Class Action Plaintiffs have served 98 document requests, 15 Interrogatories, and 939 RFAs. *Id.* ¶4. The New York Times has served 130 document requests, and 10 Interrogatories. *Id.* ¶5. And within days of each other, the *New York Times* and *Daily News* Plaintiffs served separate sets of 20 identical RFAs, nearly all of which requested admissions "with respect to each Asserted Work"—the equivalent of approximately 500 million requests. *Id.*

2832519

¶5. The *Daily News*, *CIR*, and *Raw Story* Plaintiffs have separately served similar volumes of discovery requests in each of their own cases, and all but two of these plaintiffs (The New York Times and Daily News) are separately pursuing source code and training data inspection with no coordination among themselves. *Id.* ¶6. The overwhelming majority of these discovery and inspection requests cover the same subject matter.

Despite extensive negotiations in recent months, it has become clear that plaintiffs will not—or cannot—coordinate discovery amongst themselves. *See, e.g., Authors Guild*, Dkt. 81 at 7 (describing failed and inadequate early attempts by class counsel to communicate about potential coordination). On May 17, 2024, OpenAI requested a conference with Judge Stein to discuss the need for discovery coordination between, at least, the New York and California class actions. *Id.* Dkt. 150. On May 22, counsel for OpenAI convened a meet-and-confer with counsel for the New York Times, the Daily News Plaintiffs, the New York Class Plaintiffs, and the California Class Plaintiffs to discuss deposition coordination—at which every New York plaintiff made clear that they would not coordinate with the California Class Plaintiffs. Ybarra Decl. Ex. 5. And on May 23, 2024, OpenAI sought a protective order in California to stay an early deposition of an OpenAI employee whose deposition was being simultaneously sought in both the New York and California class actions. *In re ChatGPT*, Dkt. 143. The next day, Magistrate Judge Robert Illman, to whom discovery disputes in the California Consolidated Class Action have been assigned, granted OpenAI's motion for protective order in part, recognizing the need for "all counsel in the OpenAI cases [to] meet and confer further such as to explore every avenue through which the discovery process (and depositions in particular) in these cases may be streamlined and made efficient." *Id.* Dkt. 144 at 3.

With this guidance from Judge Illman, OpenAI undertook considerable efforts over the ensuing months to negotiate a deposition coordination protocol among the California and New York Consolidated Class Actions and the Consolidated News Actions. OpenAI immediately sought to schedule the meet-and-confers ordered by Judge Illman—despite being initially shut out from those conversations by plaintiffs—and over the next several months attempted to push

10

2832519

those conversations forward, making repeated concessions in an effort to achieve some degree of reasonable coordination. Ybarra Decl. Ex. 6. OpenAI proposed a deposition coordination protocol that would truly coordinate depositions, ensuring that witnesses were not deposed more than once on duplicative, overlapping topics while allowing more than sufficient deposition time to provide each plaintiff with the discovery they need. *Authors Guild*, Dkt. 216-1 (chart comparing Defendants' and plaintiffs' proposals).

The News and Class Plaintiffs were unwilling to agree; indeed, they could not even agree to a single counterproposal and instead presented three competing proposals. *Id.* Each of these proposals, however, called for multiple depositions of each witness, agreeing only that witnesses would not be deposed more than once across the two Class Cases or across the Consolidated News Cases. *Id.* But even then, the plaintiffs insisted that an individual deponent also designated as a 30(b)(6) witness could be deposed a second time. *Id.* In other words, many of OpenAI's witnesses would have to sit for deposition *four times* just in the Consolidated News Action and the two Consolidated Class Actions. And those numbers do not account for the *Intercept, Millette*, or *Raw Story* cases, where many of the same witnesses are likely to be deposed.

Because the parties could not reach agreement on a coordination protocol, OpenAI requested that Judge Wang enter its proposed protocol, and the New York Class and News Plaintiffs requested entry of their own. During an October 30, 2024, hearing at which Judge Wang heard argument on the parties' competing deposition protocols, she stated that she was "not going to require anybody to formally coordinate with the actions in the Northern District of California" and that she "will follow Rule 30" and "will not presumptively limit witness depositions." Ybarra Decl. Ex. 1 (Oct. 30, 2024 Hearing Tr., *Authors Guild*) at 6:25-7:2; 9:7-8. The next day, Judge Wang denied all parties' request for entry of a deposition protocol. *Authors Guild*, Dkt. 252.

As of the date of this filing, the parties are in the midst of document production in the California Consolidated Class Action, the New York Consolidated Class Action, and the Consolidated News Actions. Document production was ongoing in *Raw Story*, but has been

2832519

paused while Plaintiffs seek leave to amend. No depositions have occurred in any of the copyright actions filed against OpenAI, but the parties expect—and some plaintiffs have expressly requested—that depositions will begin in the first quarter of 2025. As of the date of this filing, there is no order or stipulation coordinating depositions in any capacity in any OpenAI copyright action. Because of the unsuccessful attempts to negotiate informal discovery coordination, Judge Wang's denial of OpenAI's request for entry of a deposition coordination protocol, and emerging inconsistent decisions on substantive questions of law, OpenAI now moves for centralization under Section 1407.

III.   ARGUMENT

A.   Transfer of the OpenAI Copyright Actions for Pre-trial Coordination and Consolidation is Appropriate Under 28 U.S.C. § 1407

Transfer and consolidation under 28 U.S.C. § 1407 is warranted where (1) the cases involve one or more common questions of fact; and (2) transfer will serve the convenience of the parties and witnesses and promote the just and efficient conduct of the cases. 28 U.S.C. § 1407. Both factors are satisfied here.

1.   These actions involve numerous common questions of fact.

Each of the eight actions pending against OpenAI arises from the same or similar alleged conduct—OpenAI's purported use of plaintiffs' copyrighted works to train the LLMs powering ChatGPT.[7] Each operative complaint contains these core allegations:

Case	Allegations
In Re ChatGPT	OpenAI "harvest[ed] mass quantities of textual material from the public internet, including Plaintiffs' books," to "train the OpenAI Large Language Models[.]" Dkt. 120 ¶ 65.
Authors Guild	OpenAI "fed [authors'] copyrighted works] into their 'large language models' or 'LLMS'[.]" Dkt. 69 ¶ 2.

---

[7] For the avoidance of doubt, OpenAI does not concede that the overlapping nature of the factual allegations makes class treatment appropriate in any of the putative class actions.

2832519

NY Times	OpenAI "copied [a] mass of Times copyrighted content . . . for the purpose of 'training' [OpenAI's] GPT models." Dkt. 1 ¶ 92.
Daily News	OpenAI "scrap[ed] the newspapers' content . . . [to] train their [generative AI] systems." Dkt. 1 ¶ 13.
CIR	OpenAI "populated [its] training sets with works of journalism[.]" Dkt. 1 ¶ 10.
Raw Story	OpenAI "populated [its] training sets with works of journalism[.]" Dkt. 1 ¶ 6.
Intercept	OpenAI "populated [its] training sets with works of journalism[.]" Dkt. 87 ¶ 6.
Millette	OpenAI "transcrib[ed] YouTube videos to create training datasets that they then use to train their AI products" Dkt. 47 ¶ 4.

The overlapping factual issues are also uniquely complex and made more so by OpenAI's common defense across all cases that training LLMs constitutes transformative fair use. This complexity weighs in favor of centralization. For example, in *In re Fisher-Price Rock 'N Play Sleeper Mktg., Sales Pracs., & Prod. Liab. Litig.*, this Panel ordered centralization despite the plaintiffs' protestations that the factual issues and claims in the false advertising litigation around a defective baby product were not sufficiently complex. 412 F. Supp. 3d 1357, 1359 (J.P.M.L. 2019). The Panel disagreed, reasoning that "discovery regarding [the design and risks of the product] will be more akin to the discovery required in products liability litigations than run-of-the-mill false advertising lawsuits, with expert testimony required on a number of complicated topics." *Id.* The same is true here. These are not "run-of-the-mill" copyright lawsuits. Rather, each action will require discovery into and expert analysis of the training and design of LLMs— complex, novel and "complicated" technology. Indeed, OpenAI has made available for inspection several hundred terabytes of training data—the equivalent of approximately 140 billion pages—in addition to large excerpts of extremely complex and sensitive source code and the relevant LLMs themselves. To interpret all of this complex discovery, the California Class

2832519

Plaintiffs have disclosed 10 experts, the Daily News Plaintiffs have disclosed 14 experts, and the New York Times has disclosed *22 experts*.[8] The complexity of these overlapping factual issues amplifies the need for Section 1407 centralization of this litigation.

The asserted causes of action similarly overlap. All eight cases assert rights stemming from the same alleged violations of copyright.[9] Six cases assert direct copyright infringement claims; four raise contributory copyright infringement claims; three raise vicarious copyright infringement claims; five assert DMCA claims; four bring unfair competition claims; and two raise trademark dilution claims. Any minor differences in the exact causes of action are "not significant where … the actions still arise from a common factual core." *In re Marriott Int'l, Inc., Customer Sec. Breach Litig.*, 363 F. Supp. 3d 1372, 1374 (J.P.M.L. 2019); *In re Fisher-Price*, 412 F. Supp. 3d 1357, 1359 (holding that false advertising, fraud, unjust enrichment, and consumer protection claims were sufficiently similar to warrant centralization).

> **2.   Centralization will serve the convenience of the parties and witnesses and promote the just and efficient conduct of the actions.**

Section 1407 centralization is necessary to limit duplicative discovery. Limiting duplicative discovery is among the foremost goals of Section 1407 centralization. Manual for Complex Litigation, Fourth § 20.131 (2004) ("The objective of transfer [through the MDL process] is to eliminate duplication in discovery, avoid conflicting rulings and schedules, reduce litigation cost, and save the time and effort of the parties, the attorneys, the witnesses, and the courts."). Of potentially duplicative forms of discovery, depositions "tend to be the most costly and time-consuming activity in complex litigation" and courts should "manage the litigation so as to avoid unnecessary depositions, limit the number and length of those that are taken, and ensure that the process of taking depositions is as fair and efficient as possible." *Id.* § 11.45.

---

[8] The New York Class Plaintiffs are not required to disclose their experts under their protective order and other plaintiffs have not yet disclosed experts.

[9] The only two cases that do not explicitly plead Copyright Act infringement claims are *Raw Story* and *The Intercept*. But both of these cases effectively assert copyright infringement

2832519

The risk that the same witnesses will be deposed in each action is stark: to-date, there are already 22 current and former OpenAI employees who have been separately designated as custodians in the Consolidated News Action, the New York Class Action and the California Class Action[10]—in addition to many others for whom OpenAI has agreed to cross-produce documents without making those employees full custodians. Ybarra Decl. ¶8. And plaintiffs continue to demand more; possibly enough to double the number of overlapping custodians. Given the breadth of document requests, there is little doubt that multiple plaintiff groups will seek the depositions of these witnesses. Absent centralization, these witnesses (as well as many non-party witnesses) are virtually guaranteed to sit for multiple depositions, whether as individuals or 30(b)(6) designees.

The extensive overlap in OpenAI custodians underscores the factual similarities between each pending action. Even the plaintiffs have recognized this. For example, the California Class Action Plaintiffs recently moved to add six additional OpenAI custodians also sought in the New York Consolidated Class Action. *In re ChatGPT*, Dkt. 204. They offered to drop that motion only if OpenAI "agree[d] to add any custodians added in the *Authors Guild* Action," arguing that such reciprocal addition of custodians is a "common sense efficient approach" consistent with Judge Illman's call for the parties to "streamline" discovery. *Id*. Yet these same Plaintiffs ignore Judge Illman's guidance in refusing to coordinate depositions. The New York plaintiffs have taken the same approach. However, the New York Magistrate Judge has stated that she will not order formal cross-district deposition coordination and, to-date, has declined to order deposition coordination even in the cases pending before her. Ybarra Decl. Ex. 1 at 6:25-7:2.

All parties recognize the duplicative discovery at issue and all parties recognize the need to streamline it. But in light of plaintiffs' refusal to agree to single depositions of witnesses and the New York Magistrate Judge's recent denial of OpenAI's request for cross-district coordination, the only viable means of doing so is through Section 1407 centralization.

---

[10] Thirteen of these OpenAI custodians have also been designated in *Raw Story*.

2832519

Section 1407 centralization will avoid inconsistent rulings on important and emerging questions of law. This Panel often orders centralization where there is a risk of inconsistent pretrial rulings on important merits or procedural issues. *See, e.g., In re Auryxia (Ferric Citrate) Pat. Litig.*, 412 F. Supp. 3d 1347, 1349 (J.P.M.L. 2019) (centralizing four actions across two districts to prevent inconsistent rulings on claim construction and patent validity). This risk is especially acute in the OpenAI Copyright Litigation.

It is not hyperbole to state that the dispositive motions in these copyright cases— including especially those relating to OpenAI's fair use defense—may chart the course for artificial intelligence development in coming years. But without consolidation, two judges within the Northern District of California and three separate judges within the Southern District of New York will each have to make separate inquiries into a nearly identical—but highly complex—set of facts about the training of OpenAI's LLMs and the transformative nature of that process, and then make separate conclusions about the legal import of those facts. These inquiries will arise in the context of everything from discovery motions to summary judgment, class certification, *Dauberts* and other motions *in limine*. Each of these rulings will create the potential for inconsistent guidance on important and emerging issues of law. These inconsistencies will not only hamper and prejudice the parties in this litigation, but risk creating confusion and uncertainty in the development of LLMs and artificial intelligence more broadly, stymying growth and innovation in a burgeoning field.

Fractures are already emerging in early rulings on merits, discovery, and procedural issues. For example, while Magistrate Judge Illman has directed the California Consolidated Class Action to work to coordinate depositions with the New York Consolidated Class Action, Magistrate Judge Wang has stated that she will not order formal deposition coordination between the two cases. And recently, Judge McMahon and Judge Rakoff issued conflicting orders on near-identical issues in OpenAI's motions to dismiss the *Raw Story* and *Intercept* complaints. Judge McMahon ruled that *Raw Story* had insufficiently alleged concrete harm arising from OpenAI's alleged removal of CMI from Raw Story's works during the LLM training process,

16

and thus lacked standing to bring a DMCA claim. *Raw Story*, Dkt. 117. Judge Rakoff, on the other hand, allowed Intercept's DMCA claim to survive, notwithstanding near identical deficiencies in alleged harm. *Intercept*, Dkt. 122. In other words, two judges have reached different outcomes on the same legal issue presented by two near-identical complaints brought by the same counsel, in the same district. Section 1407 centralization is intended to prevent precisely such inconsistencies. These fractures will only grow as the litigation progresses. This Panel can ensure that the courts issue consistent guidance in this vanguard litigation by centralizing the litigation in a single district, before a single judge.

     Transfer and consolidation will avoid overlapping classes. The "potential for conflicting or overlapping class actions presents one of the strongest reasons for [centralizing] such related actions." *In re Plumbing Fixtures*, 308 F. Supp. 242, 244 (J.P.M.L. 1970). Indeed, where there is a risk of overlapping classes, this Panel has deemed it "necessary" to centralize all cases with a single judge, even in instances where there are considerable non-overlapping factual issues. *See In re Republic Nat'l-Realty Equities Sec. Litig.*, 382 F. Supp. 1403, 1406 (J.P.M.L. 1974) ("[I]t is necessary to have all class action questions resolved by a single judge.").

     As with the risk of inconsistent rulings, the risk of conflicting or overlapping classes is squarely presented in this litigation. The California Consolidated Class Action seeks to represent a class of ***all*** United States copyright holders whose works were used to train OpenAI's LLMs. The New York Consolidated Class Action seeks relief on behalf of separately defined classes of fiction and non-fiction book authors whose works were used to train OpenAI's LLMs. And the *Millette* Plaintiffs seek to represent a class of YouTube creators whose copyrighted works were used to train OpenAI's LLMs. Accordingly, the entirety of the putative classes in the New York Consolidated Class Action and *Millette* action–in addition to each individual News Plaintiff–are subsumed by the broader putative class in the California Consolidated Class Action—a fact that the California Class Plaintiffs have expressly acknowledged. *In re ChatGPT*, Dkt. 98 at 6 (*The New York Times* and New York Class Actions "are either identical to or are entirely subsumed by the proposed class defined in the first-filed Tremblay Action pending in this Court"). Such

17

substantial overlap between the three putative class actions, especially when combined with the duplicative discovery likely in this litigation, is more than sufficient to support Section 1407 centralization. *See, e.g., In re Republic Nat'l-Realty Equities Sec. Litig.*, 382 F. Supp. at 1406; *In re H & R Block Mortg. Corp. Prescreening Litig.*, 435 F. Supp. 2d 1347, 1349 (J.P.M.L. 2006).

Now is the time for transfer and consolidation. In January of 2024, OpenAI agreed to not seek transfer of *The New York Times* and *Authors Guild* cases under Section 1404, while reserving the right to seek consolidation and transfer under Section 1407. At the time, OpenAI was optimistic that it could coordinate discovery either informally among the three existing plaintiffs' groups, or through court order. OpenAI has since negotiated extensively with plaintiffs towards that goal. Over the past six months, however, as the number of cases and plaintiffs' groups have multiplied, and with the New York Magistrate Judge declining OpenAI's request for formal cross-district deposition coordination, the prospect of coordination or further consolidation absent Section 1407 centralization has dimmed.

Efforts by plaintiffs' groups to eliminate the multi-district (and multi-judge) character of the litigation have been unsuccessful. The California Class Action Plaintiffs' motions to intervene in and dismiss or, in the alternative, transfer the New York Consolidated Class Action and *New York Times* actions under the first-to-file rule were denied in each case. *Authors Guild*, Dkt. 71; *NY Times*, Dkt. 47. As were the *Raw Story* and *Intercept* Plaintiffs' requests to relate their cases to the *New York Times* action. Those cases are now assigned to separate judges within the same district, who also declined to relate the cases to each other. OpenAI's efforts to informally coordinate discovery or obtain a coordination order have also thus far not succeeded.

Meanwhile, since the January 2024 stipulation, five additional cases have been filed against OpenAI, compounding the sprawling nature of this litigation and rendering further efforts at discovery coordination impractical in light of the large number of plaintiffs and groups of counsel. *See In re Uber Technologies, Inc., Data Security Breach Litigation*, 304 F. Supp. 3d 1351 (J.P.M.L. 2018) (informal coordination impractical in multidistrict litigation involving 14 groups of plaintiffs' counsel); *In re Fisher-Price*, 412 F. Supp. 3d at 1359 (informal discovery

18

impracticable among five groups of plaintiffs' counsel). In sum, while the parties have obtained limited consolidation of the individual actions, the current multi-district, multi-judge nature of the litigation is unlikely to change going forward.

OpenAI reserved the right to seek a Section 1407 transfer in the January 2024 stipulation to guard against these exact circumstances—the filing of more cases and insufficient coordination among them. The intervening months between that stipulation and this filing have firmly established the need for centralization. And the cases remain in sufficiently similar procedural posture that Section 1407 makes sense now. While some cases have progressed further in document discovery than others, modest differences in procedural posture should not be a barrier to Section 1407 centralization where (1) no case has proceeded to depositions; (2) no expert reports have been exchanged; (3) no class certification motion has been filed; and (4) no summary judgment motion has been filed. There are many efficiencies to be obtained from centralization, and each of the considerations above counsel in favor of it.

B.   The Northern District of California is the Superior Transferee Forum

Transfer to and consolidation in the Northern District of California would make this litigation less costly and more convenient for all parties. Not only were the first cases filed in the Northern District of California, but the overwhelming majority of relevant documents are located there as well. The focus of all claims across these cases is the design and training of OpenAI's LLMs. These activities all occurred at OpenAI's headquarters in San Francisco[11] Indeed, source code and training data inspections have all occurred in the Northern District of California. Ybarra Decl. ¶7. And of the current OpenAI employees requested as custodians—the witnesses most likely to be deposed--all but two are based on the Northern District of California. *Id.* ¶9 The figures are likely the same for former OpenAI employees. A significant portion of third party

---

[11] Plaintiffs' claims (in five of eight cases) against co-defendant Microsoft also stem from Microsoft's alleged use of OpenAI's products. *See, e.g., NY Times,* Dkt. 1 ¶ 6 (claims against OpenAI and Microsoft are "deeply intertwined."). Thus, the evidence and witnesses relevant to those claims are also primarily based in the Northern District of California.

2832519

discovery will also take place in the Northern District of California. For example, on December 3, 2024, the New York Times served subpoenas on three AI companies, all based in the Northern District of California: Anthropic, PBC, Particle.News, and Perplexity AI, Inc. Ybarra Decl. ¶10. This convergence of evidence and witnesses in the Northern District of California makes it the most appropriate forum for the actions. *See In re Hard Disk Drive Suspension Assemblies Antitrust Litig.*, 396 F. Supp. 3d 1374, 1375 (J.P.M.L. 2019) (transferring actions to the Northern District of California because "common documents and witnesses likely will be located" there).

The Northern District of California also has a proven track record of efficient case management. *In re Am. Med. Collection Agency, Inc., Customer Data Sec. Breach Litig.*, 410 F. Supp. 3d 1350, 1354 (J.P.M.L. 2019) (selecting a transferee judge "with the ability and willingness to manage this litigation efficiently."). For example, in the Northern District of California, the median time to resolve a motion to dismiss is 70 days from the date of the opening motion; for summary judgment motions, the median time is 78 days. Ybarra Decl. Ex. 7 at 3. In the Southern District of New York, the median time to resolve those same motions is 161 and 200 days, respectively. Ybarra Decl. Ex. 8 at 4. And the median time to trial in the Northern District of California is 778 days. In the Southern District of New York, it is 906 days. Ybarra Decl. Ex. 9 at 3. In sum, while both courts are highly capable, the Northern District of California is better suited to handle this case because most witnesses and evidence are located here, and because of its strong track record for efficient case management.

IV.   CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that this Panel transfer all actions in Schedule A to the Northern District of California for pre-trial consolidation.

2832519

Respectfully submitted,

Dated: December 6, 2024

KEKER, VAN NEST & PETERS LLP

By: */s/ Michelle S. Ybarra*
ROBERT A. VAN NEST
RVanNest@keker.com
R. JAMES SLAUGHTER
RSlaughter@keker.com
PAVEN MALHOTRA
PMalhotra@keker.com
MICHELLE S. YBARRA
MYbarra@keker.com
NICHOLAS S. GOLDBERG
NGoldberg@keker.com
THOMAS E. GORMAN
TGorman@keker.com
KATIE LYNN JOYCE
KJoyce@keker.com
CHRISTOPHER S. SUN
CSun@keker.com
KEKER, VAN NEST & PETERS LLP
633 Battery Street
San Francisco, California 94111-1809
Telephone: (415) 391-5400
Facsimile: (415) 397-7188

21

2832519

Dated: December 6, 2024

MORRISON & FOERSTER LLP

By: */s/ Tiffany Cheung*
JOSEPH C. GRATZ
JGratz@mofo.com
TIFFANY CHEUNG
TCheung@mofo.com
VERA RANIERI
VRanieri@mofo.com
JOYCE C. LI
JoyceLi@mofo.com
MELODY E. WONG
MelodyWong@mofo.com
MORRISON & FOERSTER LLP
425 Market Street
San Francisco, California 94105-2482
Telephone: (415) 268-7000
Facsimile: (415) 268-7522

ROSE S. LEE
RoseLee@mofo.com
ALEXANDRA M. WARD
AlexandraWard@mofo.com
MORRISON & FOERSTER LLP
707 Wilshire Boulevard
Los Angeles, California 90017-3543
Telephone: (213) 892-5200
Facsimile: (213) 892-5454

CAROLYN M. HOMER
cmhomer@mofo.com
MORRISON & FOERSTER LLP
2100 L Street, NW
Suite 900
Washington, DC 2003
Telephone: 202-65004597

JOHN R. LANHAM
jlanham@mofo.com
MORRISON & FOERSTER LLP
12531 High Bluff Dr.
Suite 100
San Diego, CA 921320
Telephone 858-720-5100

MAX I. LEVY
MLevy@mofo.com
MORRISON & FOERSTER LLP
755 Page Mill Road
Palo Alto, California 94304-1018
Telephone: (650) 813-5600

22

2832519

ERIC K. NIKOLAIDES (*pro hac vice*)
ENikolaides@mofo.com
MORRISON & FOERSTER LLP
250 West 55th Street
New York, NY 10019-9601
Telephone: (212) 468-8000

Dated: December 6, 2024

LATHAM & WATKINS LLP

By: */s/Elana Nightingale-Dawson*
ANDREW M. GASS
Andrew.Gass@lw.com
JOSEPH R. WETZEL
Joe.Wetzel@lw.com
LATHAM & WATKINS LLP
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: (415) 391-0600

SARANG VIJAY DAMLE (*pro hac vice*)
Sy.Damle@lw.com
ELANA NIGHTINGALE DAWSON (*pro hac vice*)
Elana.nightingaledawson@lw.com
MICHAEL A. DAVID (*pro hac vice*)
michael.david@lw.com
LATHAM & WATKINS LLP
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: (202) 637-2200

ALLISON L. STILLMAN (*pro hac vice*)
Alli.Stillman@lw.com
RACHEL R. BLITZER (*pro hac vice*)
Rachel.blitzer@lw.com
HERMAN H. YUE (*pro hac vice*)
Herman.yue@lw.com
LATHAM & WATKINS LLP
1271 Avenue of the Americas
New York, NY 10020
Telephone: (212) 751-4864

Attorneys for Defendants
OPENAI, INC., OPENAI, L.P., OPENAI
OPCO, L.L.C., OPENAI GP, L.L.C., OPENAI
STARTUP FUND GP I, L.L.C., OPENAI
STARTUP FUND I, L.P., AND OPENAI
STARTUP FUND MANAGEMENT, LLC

2832519