U.S. DISTRICT COURT
APR 21 2025
CLERK'S OFFICE

UNITED STATES JUDICIAL PANEL
on
MULTIDISTRICT LITIGATION

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION

MDL No. 3143

TRANSFER ORDER

Before the Panel:* The OpenAI Defendants[1] move under 28 U.S.C. § 1407 to centralize this litigation in the Northern District of California. This litigation consists of twelve actions pending in the Northern District of California and the Southern District of New York, as listed on Schedule A.

Microsoft Corporation, a defendant in several of the New York actions, supports centralization in either the Northern District of California or the Southern District of New York, and asks that the MDL be assigned to an experienced transferee judge. Plaintiffs in the *Millette* action support centralization in the Northern District of California. All other responding plaintiffs oppose centralization, though their alternative positions differ. Plaintiffs in the California class actions (*Tremblay*, *Silverman*, and *Chabon*) alternatively support the Northern District of California as the transferee district. Plaintiffs in the two New York class actions (*Authors Guild* and *Alter*) alternatively suggest the Southern District of New York as the transferee district. Plaintiffs in the New York "News" cases (*New York Times*, *Daily News*, and *Center for Investigative Reporting*) and the New York "DMCA" cases (*Raw Story Media* and *The Intercept*) alternatively suggest that, if an MDL is created, the News and DMCA cases should be centralized in the Southern District of New York.[2]

On the basis of the papers filed and the hearing session held, we find that the actions listed on Schedule A involve common questions of fact, and that centralization in the Southern District of New York will serve the convenience of the parties and witnesses and promote the just and

---

* Judge David C. Norton did not participate in the decision of this matter.

[1] OpenAI, Inc.; OpenAI, L.P.; OpenAI OpCo LLC; OpenAI GP LLC; OpenAI LLC; OpenAI Global LLC; OAI Corporation LLC; OpenAI Holdings LLC; OpenAI Startup Fund I LLC; OpenAI Startup Fund GP I LLC; and OpenAI Startup Fund Management LLC (collectively, OpenAI).

[2] This alternative suggestion is a nonstarter because the News cases and the DMCA cases are already pending in the Southern District of New York and thus lack the multidistrict character required by 28 U.S.C. § 1407(a).

efficient conduct of this litigation. These actions share factual questions arising from allegations that OpenAI and Microsoft used copyrighted works, without consent or compensation, to train their large language models (LLMs), such as GPT-4, which underlie defendants' generative artificial intelligence products, such as OpenAI's ChatGPT and Microsoft's Bing Chat (rebranded as Copilot), which can algorithmically simulate human reasoning and inference.[3] Each action will involve overlapping, complex, and voluminous discovery regarding how defendants trained and designed their LLMs.[4] Given the novel and complicated nature of the technology, there likely will be overlapping experts across these actions. Centralization will eliminate duplicative discovery; prevent inconsistent pretrial rulings, particularly as to class certification; and conserve the resources of the parties, their counsel, and the judiciary.

The opposing plaintiffs focus on the differences among the claims in these actions. For instance, they contend class action cases are brought by authors of fiction and non-fiction works (primarily novels) who allege that their copyrighted works were infringed when defendants used them as *input* in the training of their LLMs whereas the News plaintiffs assert that millions of their copyrighted works (newspaper articles) were infringed both when used as input in the training of defendants' LLMs and by the LLMs *output*—specifically, that defendants' artificial intelligence products generate verbatim and detailed summaries of news content. Plaintiffs also argue that the DMCA plaintiffs, unlike plaintiffs in the other cases, do not allege copyright infringement. Instead, they allege that defendants violated the Digital Millennium Copyright Act (DMCA), 17 U.S.C. § 1202, which, *inter alia*, prohibits the removal of copyright management information (such as author, title, copyright, and terms of use information) from a work. Finally, plaintiffs in the *Millette* action, unlike the other plaintiffs, seek to represent a putative class of YouTube content creators whose videos (or, more specifically, transcripts of those videos) were used to train defendants' LLMs.

These differences in claims and the underlying material alleged to be infringed do not present a significant obstacle to centralization given the substantial overlap in factual questions

---

[3] According to the parties, LLMs are algorithms that predict words that are likely to follow a given string of text based on the potentially billions of examples used to train it. LLMs encode the information from a training corpus that they use to make these predictions as numbers called "parameters." There are, for example, approximately 1.76 trillion parameters in the GPT-4 LLM. Training an LLM allegedly involves storing encoded copies of the training works in computer memory, repeatedly passing them through the model with words masked out, and adjusting the parameters to minimize the difference between the masked-out words and the words the model predicts to fill in. After being trained on a general corpus, LLMs may be subject to "fine-tuning" by, for example, performing additional rounds of training using specific types of works to better mimic their content or style, or providing them with human feedback to reinforce desired or suppress undesired behaviors. Once trained, LLMs generate output based on patterns and connections drawn from the training data.

[4] According to the parties, defendants have already produced petabytes of data from their training sets.

and discovery relating to defendants' training of their LLMs.[5] It is not uncommon for transferee courts to establish separate tracks for actions involving differing claims that allow the actions to progress efficiently. *See, e.g.*, *In re Jan. 2021 Short Squeeze Trading Litig.*, MDL No. 2989, 2021 WL 1258399, at *3 (J.P.M.L. Apr. 2, 2021) ("The transferee court, though, can employ any number of pretrial techniques—such as establishing claim-specific or defendant specific tracks and creating an attorney leadership structure that reflects the differences in the claims—to manage the differences that these actions may present.").

The opposing plaintiffs also argue that Judge Araceli Martínez-Olguín in the Northern District of California and Judge Sidney H. Stein in the Southern District of New York, as well as their assigned magistrate judges, are already closely coordinating with one another with respect to discovery. While this informal coordination is commendable, we are persuaded that Section 1407 centralization will ensure overall economies. The currently assigned judges and magistrate judges have dedicated substantial time and resources to this litigation, and far more will be required as pretrial proceedings progress to the deposition stage. Moreover, these judges will be called upon to resolve duplicative discovery disputes and overlapping dispositive and class certification motions. Centralization will allow a single judge to coordinate discovery, streamline pretrial proceedings, and eliminate inconsistent rulings.

The opposing plaintiffs insist that pretrial proceedings in the class actions and the News actions are too advanced for centralization to be beneficial. We do not find this argument persuasive. Counsel stated during oral argument that only four depositions have been taken thus far and, at a minimum, forty more depositions will be needed in the class actions alone. The parties have been unable to agree on a deposition coordination protocol to eliminate duplicative depositions of defense witnesses—indeed, much of the recent briefing by the parties and oral argument before the Panel consisted of counsel blaming the other side for the breakdown in negotiations. Opportunities abound for the transferee judge to redirect the parties onto a more efficient path and eliminate the potential for duplicative discovery and pretrial motion practice, as well as inconsistent pretrial rulings and scheduling.

Finally, the opposing plaintiffs argue that OpenAI cannot seek centralization under the terms of a January 2024 stipulation and order in the New York class actions. We need not resolve the parties' dispute regarding the interpretation of this stipulation because, regardless of whether OpenAI could request centralization under the stipulation, this Panel itself can certainly order it. *See* 28 U.S.C. § 1407(c) (granting the Panel the authority to centralize civil actions upon its own initiative); Hearing Session Order, MDL No. 3143 (J.P.M.L. Feb. 14, 2025), ECF No. 65 ("IT IS FURTHER ORDERED that the Panel may, on its own initiative, consider transfer of any or all of the actions in those matters [listed on the attached Schedule] to any district or districts."). *Cf. Uber Techs., Inc. v. U.S. Jud. Panel on Multidistrict Litig.*, No. 23-3445, __ F.4th __, 2025 WL 748135, at *8 (9th Cir. Mar. 10, 2025) ("It is clear from the text of Section 1407 that the statute does not create an individual right to centralization that may be waived but instead vests the JPML with a

[5] Even the most unique of these actions—the *Millette* YouTube action—will share common factual questions and significant discovery regarding the training of defendants' LLMs.

power to manage the federal docket by centralizing cases that is unfettered by private agreements.").

The Southern District of New York is an appropriate transferee district for this litigation. Eight of the twelve actions in this docket are already pending in this convenient and accessible venue. Several plaintiffs, as well as Microsoft Corporation, alternatively suggest or do not oppose centralization in the Southern District of New York. Judge Sidney H. Stein, an experienced transferee judge, presides over six of the eight actions in the district. Both he and Magistrate Judge Ona T. Wang have devoted substantial time and resources to this litigation, and thus are well situated to steer this litigation on a prudent and expeditious course.

IT IS THEREFORE ORDERED that the actions listed on Schedule A and pending outside the Southern District of New York are transferred to the Southern District of New York and, with the consent of that court, assigned to the Honorable Sidney H. Stein for coordinated or consolidated pretrial proceedings.

PANEL ON MULTIDISTRICT LITIGATION

Karen K. Caldwell
Chair

Nathaniel M. Gorton
Roger T. Benitez
Madeline Cox Arleo

Matthew F. Kennelly
Dale A. Kimball

A CERTIFIED COPY
TAMMI M. HELLWIG, CLERK
BY
Deputy Clerk

IN RE: OPENAI, INC., COPYRIGHT INFRINGEMENT LITIGATION MDL No. 3143

SCHEDULE A

Northern District of California

TREMBLAY, ET AL. v. OPENAI, INC., ET AL., C.A. No. 3:23−03223
SILVERMAN, ET AL. v. OPENAI, INC., ET AL., C.A. No. 3:23−03416
CHABON, ET AL. v. OPENAI, INC., ET AL., C.A. No. 3:23−04625
MILLETTE v. OPENAI, INC., ET AL., C.A. No. 5:24−04710

Southern District of New York

AUTHORS GUILD, ET AL. v. OPENAI, INC., ET AL., C.A. No. 1:23−08292
ALTER, ET AL. v. OPENAI, INC., ET AL., C.A. No. 1:23−10211
THE NEW YORK TIMES COMPANY v. MICROSOFT CORPORATION, ET AL., C.A. No. 1:23−11195
BASBANES, ET AL. v. MICROSOFT CORPORATION, ET AL., C.A. No. 1:24−00084
RAW STORY MEDIA, INC., ET AL. v. OPENAI, INC., ET AL., C.A No. 1:24−01514
THE INTERCEPT MEDIA, INC. v. OPENAI, INC., ET AL., C.A. No. 1:24−01515
DAILY NEWS LP, ET AL. v. MICROSOFT CORPORATION, ET AL., C.A. No. 1:24−03285
THE CENTER FOR INVESTIGATIVE REPORTING, INC. v. OPENAI, INC., ET AL., C.A. No. 1:24−04872