

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

THE INTERCEPT MEDIA, INC.,

Plaintiff,

v.

OPENAI, INC., OPENAI GP, LLC, OPENAI,
LLC, OPENAI OPKO, LLC, OPENAI GLOBAL,
LLC, OAI CORPORATION, LLC, OPENAI
HOLDINGS, LLC, and MICROSOFT
CORPORATION,

Defendants.

Civil Action No. 1:24-cv-01515-JSR

**SUPPLEMENTAL RULE 26(f) REPORT AND
PROPOSED CASE MANAGEMENT PLAN**

In addition to the Form D Civil Case Management Plan filed herewith, the parties hereby submit this supplemental report of their Rule 26(f) conference and proposed case management plan (“Supplemental Report”).

Parties’ Joint Request. All parties agree and respectfully request an alternative trial-ready date and schedule as set forth in this Supplemental Report. As described below, this case raises issues regarding training of large language machine learning models and the commercialization of technology using those models. Those issues are also presented in a number of other pending actions. Defendants believe that discovery into these issues regarding this technology presents (among other concerns) two significant technical challenges, both of which arise from the fact that the relevant data is far too voluminous for traditional ESI discovery procedures. Those challenges

are in the process of being solved in connection with multiple other pending actions. It would be just, speedy, and efficient for this Court to adopt a schedule consistent with those ongoing efforts to address this technology. Plaintiff lacks information to confirm Defendants' representations about Defendants' ESI and how best to address it, but Plaintiff has no present reason to dispute Defendants' characterizations. Plaintiff defers to Defendants' representations about the amount of time Defendants' need to comply with their discovery obligations, and would prefer the Court enter a schedule that Defendants can realistically meet, including a schedule that extends beyond the Court's five-month ready-for-trial date.

Defendants' Additional Grounds. Additionally, as set forth below, Defendants offer a further ground for requesting a schedule comparable to that of other pending actions. Defendants intend to file motions to dismiss all claims pled against them. Similar claims have been dismissed from a number of other actions. A longer schedule will permit the Court to consider the motions and allow the Court and the parties to avoid the significant burden of addressing the substantial technical discovery challenges on a compressed timetable in a case that may well not survive the pleading stage of the action. Even if the motions are denied, the case would not be significantly delayed in light of the parties' ongoing efforts to develop efficient solutions to the technical discovery issues in the parallel cases.

Plaintiff disputes these grounds, and notes that its allegations are significantly more detailed than any claims that have been dismissed elsewhere.

Parties' Jointly Requested Alternative Schedule

1. On March 27, 2024, the parties met and conferred over video pursuant to Rule 26(f). While the parties submit a proposed schedule in Form D in accordance with the Court's order that

the case be trial ready by September 4, 2024, counsel for Plaintiffs, Microsoft, and the OpenAI defendants respectfully request that the Court instead order the below alternative schedule.

Case Event	Parties' Proposed Date
First Date To Serve First Set of Requests for the Production of Documents	April 10, 2024
Date To Serve Interrogatories pursuant to Rule 33.3(a)	April 10, 2024
Joinder of additional parties	June 3, 2024
Amend pleadings without leave of Court	June 3, 2024
Deadline to Serve Requests to Admit	September 1, 2024
Close of fact discovery	October 4, 2024
Opening expert reports	October 18, 2024
Rebuttal expert reports	November 18, 2024
Deposition cut-off	December 9, 2024
Close of discovery	December 9, 2024
MSJ and <i>Daubert</i> motions	January 7, 2025
Oppositions to MSJ and <i>Daubert</i> motions	February 7, 2025
Replies ISO MSJ and <i>Daubert</i> motions	February 28, 2025
Final pretrial conference and MSJ oral argument.	[to be inserted by the Court]
Trial date	[to be inserted by the Court]

2. There are currently numerous cases underway, both in this District and in other Districts, that inquire into OpenAI's training of its GPT large language models and the commercial applications of that technology by OpenAI and, in some cases, Microsoft, including *J. Doe I et al., v. GitHub, Inc. et al.*, 4:22-cv-06823-JST (N.D. Cal.); *In Re OpenAI ChatGPT Litigation*, 3:23-

cv-03223-AMO, 3:23-cv-03416-AMO, 3:23-cv-04625-AMO (N.D. Cal.) (consolidating three class actions); *Authors Guild et al. v. OpenAI Inc. et al*, No. 1:23-cv-8292-SHS (S.D.N.Y.); *Alter et al. v. Open AI, Inc. et al.*, 1:23-cv-10211-SHS (S.D.N.Y.); *Basbanes et al. v. Microsoft Corp. et al.*, 1:24-cv-00084-SHS (S.D.N.Y.); *The New York Times Co. v. Microsoft Corp. et al.*, 1:23-cv-11195-SHS (S.D.N.Y.); and *Raw Story Media, Inc. v. OpenAI, Inc., et al.*, 1:24-cv-01514-UA (S.D.N.Y.).

3. The Parties request this proposed schedule because the complex technology here presents unique discovery issues that warrant an extended schedule. Large language models (“LLMs”) are created by feeding a huge amount of text through a machine learning algorithm called a “transformer,” which breaks the text into constituent “tokens,” then learns the statistical relationships among them. This allows large language models to generate language content in response to user prompts. LLMs can also be trained to perform other tasks, from translation, to computer coding, to ad copy, to research.

4. The large language models at issue in this case have been developed using an enormous corpus of textual data. Plaintiff’s claims rest primarily on the ability to discover facts about the contents of the training datasets. In the other pending lawsuits regarding this LLM technology, OpenAI has been working diligently for months to collect training data, create facilities and methods to enable the various plaintiffs to query these massive datasets, and to develop security and confidentiality protocols for inspection of the datasets, which OpenAI contends to be highly sensitive and proprietary, and in some cases subject to third party confidentiality obligations. In addition, the training datasets OpenAI used to teach the GPT models cannot be replicated for production because of the sheer volume of the data, and are not readily

available for searching in the ordinary course of business. Overall, the technical and security constraints for OpenAI to produce the datasets for inspection are considerable.

5. Plaintiff also makes certain allegations that may implicate the contents of model outputs when in use, and the extent to which the models allegedly regurgitate training material from Plaintiff's Works. Discovering information about the contents of such outputs is also a technical challenge. OpenAI and Microsoft have been working to understand what user data is available and whether and how it can be inspected. As of yet, no means of sampling and collecting anonymized prompt and output data has been devised for use in these cases, although efforts to develop such means are underway. The specific contents of individual user data are not readily observable by humans in the ordinary course of business.

6. The alternative schedule proposed here aligns with the schedule adopted in the three Consolidated Class Actions pending in this District before Judge Stein that also challenge this technology: *Authors Guild et al. v. OpenAI Inc. et al.*, No. 1:23-cv-8292-SHS; *Alter et al. v. Open AI, Inc. et al.*, 1:23-cv-10211-SHS; and *Basbanes et al. v. Microsoft Corp. et al.*, 1:24-cv-00084-SHS. The Consolidated Class Actions are on the fastest schedule to merits resolution of any of the cases pending in any District. The schedule in the Consolidated Class Actions, which the Parties here now request this Court also adopt, is the shortest schedule consistent with developing a fulsome record on this technology.

7. The scope of the issues required to be developed also favors the Parties' requested schedule. For instance, this case will likely require discovery into at least the following subjects: (1) identification of the universe of Plaintiff's works in suit; (2) whether Plaintiff's purported copyrighted works were actually used to train the large language models at issue; (3) whether and how the alleged inclusion of Plaintiff's works in the training data results in copying of those works;

(4) whether Plaintiff included copyright management information (“CMI”) in Plaintiff’s works; (5) whether that alleged CMI was intentionally removed with reasonable grounds to know that such removal will induce, enable, facilitate, or conceal an infringement of copyright; (6) whether Microsoft had any involvement in the conduct accused of violating the DMCA; (7) whether Defendants distributed copies of Plaintiff’s works with CMI removed; (8) whether the alleged CMI was intentionally removed from model outputs with reasonable grounds to know that such removal will induce, enable, facilitate, or conceal an infringement of copyright; (9) third-party content discussing or reporting on the same subject matter as Plaintiff’s works; (10) the availability of third-party content that is similar to or describes Plaintiff’s works; (11) and the availability or good-faith reliance on various subsidiary defenses to infringement, thereby negating any scienter in connection with the training and operation of the model, including but not limited to, fair use, substantial noninfringing uses, scènes à faire, 17 U.S.C. § 102(b), 17 U.S.C. § 512 (a)-(d), and de minimis copying. The requested schedule will allow the Parties to fully investigate these issues.

8. Further, as the witnesses and relevant documents largely overlap, coordinating discovery deadlines in this case and the Consolidated Class Actions will minimize the Defendants’ need to engage in multiple investigations, document collections, document reviews and productions. Aligning the case schedules will also allow the Defendants to coordinate witness deposition dates to minimize the burden and disruption to the witnesses.

Defendants’ Additional Grounds

Additionally, Defendants will be filing motions to dismiss all claims. Numerous plaintiffs have attempted to plead identical theories under Section 1202 in other very similar cases. And, in each and every case, those claims have either been dismissed or withdrawn, regardless of how they were pleaded. *See, e.g., Doe 1 v. GitHub, Inc.*, No. 22-cv-06823 (N.D. Cal.); *Tremblay v. OpenAI*,

Inc., No. 23-cv-03223 (N.D. Cal.); *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417 (N.D. Cal.); *Andersen v. Stability AI Ltd.*, No. 23-cv-00201 (N.D. Cal.); *Huckabee v. Meta Platforms, Inc.*, No. 23-cv-09152 (S.D.N.Y.). No doubt as a result of these rulings, some plaintiffs have now abandoned or decided not to pursue the Section 1202 liability theory. *See, e.g.*, Compl. in *Authors Guild v. OpenAI*, No. 23-cv-08292 (S.D.N.Y., filed Sept. 20, 2023); Compl. in *Alter v. OpenAI, Inc.*, No. 23-cv-10211 (S.D.N.Y., filed Nov. 21, 2023); Amended Compl. in *In re ChatGPT Litigation*, No. 23-cv-03223 (N.D. Cal., filed Mar. 13, 2024). For those reasons, Defendants respectfully request that the Court order a longer schedule to permit the Court to consider the legal sufficiency of these very similar Section 1202 claims while Defendants continue their efforts to address the substantial technical challenges arising from the nature of discovery in these cases.

Plaintiff disputes that its complaint is identical or even similarly situated to any that have been dismissed.

Dated: April 2, 2024

LOEVY & LOEVY

By: /s/ Matthew Topic

Jonathan Loevy (admitted *Pro Hac Vice*)
Michael Kanovitz (admitted *Pro Hac Vice*)
Lauren Carbajal (admitted *Pro Hac Vice*)
Stephen Stich Match (No. 5567854)
Matthew Topic (admitted *Pro Hac Vice*)
311 North Aberdeen, 3rd Floor
Chicago, IL 60607
312-243-5900 (p)
312-243-5902 (f)
jon@loevy.com
mike@loevy.com
carbajal@loevy.com
match@loevy.com
matt@loevy.com

Attorneys for Plaintiff
THE INTERCEPT MEDIA, INC.

Dated: April 2, 2024

LATHAM & WATKINS LLP

By: /s/ Joseph R. Wetzel

LATHAM & WATKINS LLP

Joseph R. Wetzel

joseph.wetzel@lw.com

Andrew M. Gass (*pro hac vice pending*)

andrew.gass@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

Attorneys for Defendants

OPENAI, INC., OPENAI GP, LLC, OPENAI,

LLC, OPENAI OPKO LLC, OPENAI GLOBAL

LLC, OAI CORPORATION, LLC, and OPENAI

HOLDINGS, LLC

MORRISON & FOERSTER LLP

By: /s/ Allyson Bennett

MORRISON & FOERSTER LLP

Joseph C. Gratz (*pro hac vice*)

JGratz@mofocom

Vera Ranieri (*pro hac vice pending*)

VRanieri@mofocom

425 Market Street

San Francisco, CA 94105
Telephone: 415.268.7000

Allyson R. Bennett (*pro hac vice*)
ABennett@mofo.com
Rose Lee (*pro hac vice pending*)
RoseLee@mofo.com
707 Wilshire Boulevard
Los Angeles, CA 90017
Telephone: 213.892.5355

Attorneys for Defendants
OPENAI, INC., OPENAI GP, LLC, OPENAI,
LLC, OPENAI OPKO LLC, OPENAI GLOBAL
LLC, OAI CORPORATION, LLC, and OPENAI
HOLDINGS, LLC

Dated: April 2, 2024

ORRICK, HERRINGTON & SUTCLIFFE LLP

By: /s/ Lisa T. Simpson

Annette L. Hurst (admitted *Pro Hac Vice*)

ahurst@orrick.com

Daniel Justice (admitted *Pro Hac Vice*)

djustice@orrick.com

The Orrick Building

405 Howard Street

San Francisco, CA 94105-2669

Telephone: 415.773.5700

Facsimile: 415.773.5759

Christopher Cariello

ccariello@orrick.com

Lisa T. Simpson

lsimpson@orrick.com

51 West 52nd Street

New York, NY 10019-6142

Telephone: 212.506.5000

Facsimile: 202.506.5151

Attorneys for Defendant

Microsoft Corporation

CERTIFICATE OF SERVICE

I hereby certify this second day of April, 2024, I caused a true and correct copy of the foregoing to be electronically filed with the Clerk of the court using the CM/ECF system which will send notification to the attorneys of record and is available for viewing and downloading.

/s/ Matthew Topic
Matthew Topic