UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE:

OPENAI, INC.,

COPYRIGHT INFRINGEMENT LITIGATION

This Document Relates To:

    All Actions

25-md-3143 (SHS) (OTW)

ORDER

SIDNEY H. STEIN, U.S. District Judge.

OpenAI's[1] objections (Dkt. Nos. 840 and 934) to Magistrate Judge Wang's discovery orders dated November 7, 2025 (Dkt. No. 734), and December 5, 2025 (Dkt. No. 910), are denied and the orders are affirmed.

I. BACKGROUND

News Plaintiffs[2] and Class Plaintiffs[3] have sought discovery of logs showing ordinary users' "conversations" with ChatGPT (*i.e.*, users' prompts and ChatGPT's outputs). In the ordinary course of its business, OpenAI has retained tens of billions of such logs. (Dkt. No. 40 at 1.)

In July 2025, News Plaintiffs moved to compel OpenAI to produce a sample of 120 million consumer ChatGPT logs. (Dkt. No. 394.) OpenAI opposed that motion, noting that "OpenAI ha[d] . . . proffered a 20 million conversation sample[,] . . . [which] is surely more than enough to conduct appropriate analyses relevant to Plaintiff's claim." (Dkt. No. 436 at 1.) OpenAI also noted that it would "de-identify" any conversation sample by removing both personally identifiable information and other

---

[1] This Order refers to defendants OpenAI, Inc., OpenAI OpCo LLC, OpenAI GP, LLC, OpenAI, LLC, OpenAI Global LLC, OAI Corporation, OpenAI Holdings, LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC collectively as "OpenAI."

[2] This Order refers to plaintiffs in Case Nos. 23-cv-11195, 24-cv-1515, 24-cv-3285, 24-cv-4872, and 25-cv-4315 collectively as "News Plaintiffs."

[3] This Order refers to plaintiffs in Case Nos. 23-cv-8292, 23-cv-10211, 25-cv-3291, 25-cv-3482, and 25-cv-3483 collectively as "Class Plaintiffs."

private information from the logs using "OpenAI's custom de-identification tool." (*Id.*) News Plaintiffs agreed to OpenAI's 20 million de-identified log sample proposal on August 11, 2025, while noting that they continued to request the larger sample. (Dkt. No. 719 at 3; Dkt. No. 719-2.)

On October 14, 2025, OpenAI informed News Plaintiffs that it would not produce the entirety of the 20 million de-identified log sample, instead proposing to run search terms across the 20 million de-identified log sample to identify conversations implicating News Plaintiffs' works, which OpenAI argued would better protect ChatGPT users' privacy interests. (Dkt. No. 864-18 at 8.) The next day, News Plaintiffs filed a motion to compel OpenAI to produce the full 20 million de-identified log sample. (Dkt. No. 656.)

Judge Wang granted News Plaintiffs' motion to compel on November 7, 2025, and ordered OpenAI to produce the entire de-identified 20 million log sample to News Plaintiffs. (Dkt. No. 734.)[4] OpenAI requested reconsideration (Dkt. No. 742), and Judge Wang denied reconsideration in an order dated December 2, 2025 (Dkt. No. 896). In sum, Judge Wang found that the entirety of the 20 million log sample—both logs that implicate News Plaintiffs' works and logs disconnected from News Plaintiffs' works—are relevant for this case because even "output logs that do not contain reproductions of News Plaintiffs' works may still be relevant to OpenAI's fair use defense." (Dkt No. 896 at 6.) She also found that ChatGPT users' "sincere" privacy interests were "one factor in the proportionality analysis" but that these privacy interests were adequately protected by (1) "reducing the total output logs subject to discovery from tens of billions to 20 million," (2) OpenAI's de-identification of the 20 million log sample, and (3) the existing protective order in this case. (*Id.* at 7–8.)

OpenAI has objected to Judge Wang's discovery rulings on the grounds that they were clearly erroneous or contrary to law because her balancing of privacy considerations with relevance was inadequate. (Dkt. No. 935 at 2.)

II. ANALYSIS

Judge Wang's rulings were neither clearly erroneous nor contrary to law. She adequately balanced ChatGPT users' privacy interests against the relevance of the documents in light of the privacy protections already in place.

---

[4] On December 5, 2025, Judge Wang ordered OpenAI to produce the 20 million de-identified log sample to Class Plaintiffs as well. (Dkt. No. 910 at 4.) OpenAI filed an objection to that order (Dkt. No. 934) and filed a consolidated memorandum in support of its objections to both orders regarding production of the 20 million de-identified log sample to News Plaintiffs and Class Plaintiffs. (*See* Dkt. No. 935 at 1 n.1.)

OpenAI's argument that Judge Wang failed to give appropriate weight to ChatGPT users' privacy interests relies primarily on *S.E.C. v. Rajaratnam*, 622 F.3d 159 (2d Cir. 2010). That reliance is misplaced. In *Rajaratnam*, the United States Court of Appeals for the Second Circuit granted a writ of mandamus reversing the district court's order that required the defendants to produce to the SEC recordings of wiretapped telephone conversations that the defendants had received in discovery in a parallel criminal case. *Id.* at 164–65. The Second Circuit's main concern was that the district court had erred by failing to "ascertain the legality of the wiretaps at issue, a critical factor in determining how to weigh the competing interests" and had failed to weigh properly "the privacy interests at stake against the SEC's right to disclosure." *Id.* at 169.
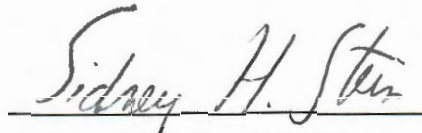
Rajaratnam is distinguishable from the issue here. First, privacy interests in the wiretapped recordings of private phone conversations in *Rajaratnam* are stronger than the privacy interests in users' conversations with ChatGPT which users voluntarily disclosed to OpenAI and which OpenAI retains in the normal course of its business. Second, *Rajaratnam* was focused on the potential illegality of the wiretaps in the first place, while no party here contends that OpenAI illegally possesses the ChatGPT conversation logs.

OpenAI also contends that it was clearly erroneous for Judge Wang to have rejected OpenAI's purportedly less burdensome proposal to run search terms across the 20 million de-identified log sample in order to identify conversations that implicate News Plaintiffs' works, which would result in the production of far fewer than 20 million conversations. Doing so, according to OpenAI, would better protect ChatGPT users' privacy interests. This argument is largely a repackaging of OpenAI's primary argument that Judge Wang failed adequately to account for users' privacy interests and should have chosen a path that would futher minimize the burden on users' privacy interests. OpenAI identifies no caselaw requiring a court to order the least burdensome discovery possible or to explain specifically why it rejects a party's discovery proposal. Judge Wang's failure to explain explicitly why she rejected OpenAI's search term proposal is not clearly erroneous or contrary to law given that she adequately explained her reasons for ordering production of the entirety of the 20 million de-identified log sample.

Accordingly, OpenAI's objections are denied and Judge Wang's orders are affirmed.

Dated: New York, New York
       January 5, 2026

SO ORDERED:

Sidney H. Stein, U.S.D.J.