UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

---x

IN RE: :

                                                            :

OPENAI, INC.,   :   25-md-3143 (SHS) (OTW)
COPYRIGHT INFRINGEMENT LITIGATION,   :

                                                            :

                                                            :   <u>**OPINION & ORDER**</u>

This Document Relates To: :
All Actions   :

                                                            :

---x

ONA T. WANG, **United States Magistrate Judge**:

The Court is in receipt of OpenAI's motion for reconsideration, (ECF 742),[1] of my

November 7, 2025, order directing OpenAI to produce the 20 million retained, de-identified

consumer ChatGPT output logs (the "20 Million ChatGPT Logs") for merits sampling purposes,

(ECF 734). For the reasons below, OpenAI's motion is **DENIED.**

I.   BACKGROUND

The Court assumes familiarity with the facts of this case and summarizes below the

factual and procedural history relevant to the pending motion.

News Plaintiffs have sought, in one form or another, output log data from OpenAI since

at least May 2024. (*See* ECF 719 at 2); (*see also* 23-CV-11195, ECF 311-2 at 2) (requesting output

logs related to New York Times content). News Plaintiffs' initial requests sought an incredibly

vast amount of data, including consumer ChatGPT output logs, enterprise ChatGPT output logs,

and Application Programming Interface ("API") output logs. (ECF 719 at 2). OpenAI has retained

tens of billions of consumer ChatGPT output logs alone in the ordinary course of its business.

---

[1] Unless otherwise indicated, all references to the docket refer to 25-MD-3143.

(ECF 40). In May 2025, News Plaintiffs learned that OpenAI had been deleting the API and enterprise ChatGPT logs, as well as a certain number of consumer ChatGPT logs that had been marked for deletion by its users. (*See* ECF 22-1 at 9). On May 13, 2025, I directed OpenAI to preserve and segregate all output log data that would have otherwise been deleted on a going forward basis until further order of the Court. (ECF 33). The Court then held several settlement conferences with the News Plaintiffs and OpenAI regarding the issue of OpenAI's deletion of consumer ChatGPT logs and potential spoliation concerns. (*See, e.g.*, ECF Nos. 217, 444, 592).[2]

During this time, the parties also began negotiating the process for sampling the tens of billions of consumer ChatGPT logs that OpenAI has retained. (*See, e.g.*, ECF 394) (seeking to compel OpenAI to produce a "statistically valid monthly sample of OpenAI's ChatGPT output logs from December 2022 through November 2024"); (ECF 435) (discussing sample sizes for a "merits sampling exercise").[3] The sampling of consumer ChatGPT logs for spoliation purposes was separate and distinct from the merits sampling exercise. (ECF 435 at 2 n. 1). On July 30, 2025, News Plaintiffs filed a motion seeking to compel OpenAI to produce a sample of 120 million consumer ChatGPT logs for its merits analysis—representing 5 million output logs per month across a 24-month period—which they assert would allow News Plaintiffs to adequately test OpenAI's various LLM models. (*See* ECF 394).

Shortly thereafter, OpenAI filed its opposition to News Plaintiffs' motion and argued that a sample size of 20 million would be "more than enough to conduct appropriate analyses

---

[2] After months of meeting and conferring, the parties filed a stipulation terminating OpenAI's ongoing preservation obligations under the Court's preservation order, (ECF 628), which was entered on October 9, 2025. (ECF 629).

[3] News Plaintiffs' motion to compel a statistically valid monthly sample of consumer ChatGPT logs for merits purposes also includes discussion regarding sampling for spoliation purposes. (23-CV-11195, ECF 788 at 2).

relevant to Plaintiffs' claims." (ECF 435). As part of its opposition, OpenAI strongly emphasized

that the increased amount of time that would be required to de-identify a sample of 120 million

output logs (36 weeks) versus a smaller sample size of 20 million (12 weeks) would not be

proportional. (*Id.* at 1). Notably, OpenAI underscored at this point that its own internal de-

identification tool is significantly more effective than the de-identification tool identified by

Plaintiffs. (*Id.*). OpenAI's representations at this point indicated that a smaller sample size of 20

million consumer ChatGPT logs and its comprehensive internal de-identification tool would

alleviate the associated privacy concerns with production of the output logs:

> Generating this sample will require OpenAI to (i) retrieve each
> requested log by finding it in the tens of billions of logs in OpenAI's
> offline data storage, (ii) decompress each log; and (iii) parse each
> log, including by sorting through layout changes that occurred
> during the Study Period. The logs must then be (iv) de-identified
> using the same rigorous process that OpenAI applies to other user
> data. That requires application of OpenAI's [internal de-
> identification tool].

(*Id.* at 2) (internal citations omitted) (emphasis added). On August 11, 2025, News Plaintiffs

agreed to proceed with the de-identification of the 20 million log sample. (ECF 719 at 3);

(*see also* ECF 719-2).

On October 14, 2025, OpenAI communicated to News Plaintiffs that it would not

produce the 20 Million ChatGPT Logs, and instead proposed to further narrow the sample by

running additional search terms across the now-de-identified logs. (ECF 656-1). News Plaintiffs

filed a motion to compel the next day, seeking an order compelling OpenAI to produce all

20 Million ChatGPT Logs. (ECF 656). OpenAI filed its opposition to News Plaintiffs' motion on

October 20, 2025. (ECF 679). In its opposition, OpenAI argues that News Plaintiffs have

somehow conceded that less than 0.1% of the 20 Million ChatGPT Logs are actually relevant,[4]

and that ordering production of the 20 Million ChatGPT Logs is not proportional under Rule 26

given the privacy concerns implicated with production. (ECF 679).

The Court heard oral argument on this issue at the October 29, 2025, discovery status

conference, where I directed the parties to submit supplemental briefing by October 30, 2025.

(ECF 715). The parties filed their supplemental briefs on October 30, 2025. (ECF Nos. 717, 718).

Then, on November 7, 2025, I granted News Plaintiffs' motion to compel and directed OpenAI

to produce the 20 Million ChatGPT Logs. (ECF 734).

OpenAI filed a letter[5] on November 12, 2025, requesting that I reconsider and vacate

my November 7 order, and that I stay the order pending resolution of OpenAI's motion. (ECF

742). That same day, News Plaintiffs filed an opposition to OpenAI's motion for reconsideration.

(ECF 746). Then, on November 13, 2025, I directed OpenAI to file a reply to News Plaintiffs'

opposition, (ECF 750), which OpenAI filed on November 14, 2025. (ECF 752).[6]

II.   LEGAL STANDARD

Reconsideration is an "extraordinary remedy to be employed sparingly in the interests

of finality and conservation of scarce judicial resources." *RST (2005) Inc. v. Research in Motion*

---

[4] In support of this assertion, OpenAI cites to News Plaintiffs' October 15, 2025, motion to compel OpenAI to produce the 20 Million ChatGPT Logs, (ECF 656), and News Plaintiffs' July 30, 2025, motion to compel OpenAI to produce a statistically valid sample of consumer ChatGPT output logs, (ECF 394), neither of which contain any "concession" by News Plaintiffs that less than 0.1% of the output logs are relevant. OpenAI goes on to say that "News Plaintiffs concede that over 99.99% of that data has no specific relevance to the claims and defenses at issue here," (ECF 679), but cites nothing in the record to support this proposition.

[5] While not styled as one, the Court construes OpenAI's letter as a letter <u>motion</u> for reconsideration.

[6] OpenAI has also filed an objection to my November 7 order. (*See* ECF Nos. 840, 841). It appears, however, that OpenAI's objection, which was filed on November 24, 2025, may be untimely. *See* Fed. R. Civ. P. 72(a) ("A party may serve and file objections to the order within 14 days after being served with a copy.").

Ltd., 597 F. Supp. 2d 362, 365 (S.D.N.Y. 2009) (quoting *In re Health Mgmt. Sys., Inc. Sec. Litig.*, 113 F. Supp. 2d 613, 614 (S.D.N.Y. 2000)). "The major grounds justifying reconsideration are 'an intervening change of controlling law, the availability of new evidence, or the need to correct a clear error or prevent manifest injustice.'" *Virgin Atl. Airways, Ltd. v. Nat'l Mediation Bd.*, 956 F.2d 1245, 1255 (2d Cir. 1992) (quoting 18 C. Wright, A Miller & E. Cooper, Federal Practice & Procedure § 4478). The movant carries the heavy burden of "point[ing] to controlling decisions or data that the court overlooked—matters in other words, that might reasonably be expected to alter the conclusion reached by the court." *Schrader v. CSX Transp., Inc.*, 70 F.3d 255, 257 (2d Cir. 1995). "A motion for reconsideration, however, is not meant to be an opportunity for the movant to relitigate old issues or take a second bite at the apple." *In re Iraq Telecom Ltd.*, 18-MC-458 (LGS) (OTW), 2019 WL 5080007, at *1 (S.D.N.Y. Oct. 10, 2019) (internal quotation marks omitted). Ultimately, whether to grant a motion for reconsideration is left to the "sound discretion of the district court." *Vincent v. Money Store*, 03-CV-2876 (JGK), 2011 WL 5977812, at *1 (S.D.N.Y. Nov. 29, 2011).

III.   DISCUSSION

As an initial matter, OpenAI's motion for reconsideration should be denied because OpenAI has failed to "put forth any facts or law that the Court did not consider and that would compel a different conclusion." *In re OpenAI, Inc., Copyright Infringement Litig.*, 25-MD-3143 (SHS) (OTW), 2025 WL 1443780, at *1 (S.D.N.Y. May 16, 2025). While the Court could conclude its analysis on this basis alone, it will address OpenAI's arguments. I find that production of the 20 Million ChatGPT Logs is both relevant and proportional.

Regardless of the applicability of the sampling methodology and subsequent production in *Concord Music Grp., Inc. v. Anthropic PBC*,[7] the 20 Million ChatGPT Logs are relevant to the case. Contrary to OpenAI's contention, News Plaintiffs have not conceded that "at least 99.99% of the conversation logs are irrelevant." (ECF 742 at 2); (*see also* ECF 746 at 2) ("OpenAI's assertion that News Plaintiffs have conceded that 99.99% of the chat outputs are irrelevant is simply false."). Moreover, the relevance of the 20 Million ChatGPT Logs is not limited to instances where News Plaintiffs' copyrighted works may be reproduced in whole or in part. Indeed, output logs that do not contain reproductions of News Plaintiffs' works may still be relevant to OpenAI's fair use defense—OpenAI has argued in the past that output logs are relevant to damages and factor four of its fair use defense, (*see* ECF 635 at 3)[8]—as well as

---

[7] 24-CV-3811 (EKL) (SVK), 2025 WL 1482734 (N.D. Cal. May 23, 2025). Despite OpenAI's contention, as discussed below, the reasoning in *Concord* was not the sole basis for the November 7 order. (*See* ECF 734). Throughout the negotiations of the sampling process, OpenAI has, repeatedly, relied on the sampling procedures in *Concord* as grounds for utilizing a sample size of 5 million logs, notwithstanding the very differences between the cases and the types and numbers of logs at issue in each that OpenAI now says make *Concord* inapplicable. (*See* May 27, 2025, Tr., ECF 291 at 23) (hereinafter, May 27 Tr.) ("And if you look at how other cases have addressed this issue, in particular, I think the most pertinent and relevant to this case is one that has just come to ground recently in the *Concord Music Group v. Anthropic* case… They've addressed this very issue about how to sample conversation data in order to figure out what is reasonable and proportionate for the purposes of the case."); (ECF 435) (OpenAI citing *Concord* as persuasive to determine the appropriate sample size). (*See also* ECF 104) (OpenAI's proposed order re conversation data sampling relying on *Concord* to determine appropriate sample size). OpenAI <u>did</u> have an opportunity to explain why a case that it has relied on consistently for nearly a year did not apply, it simply chose not to address it because it did not support OpenAI's argument.

OpenAI's remaining arguments regarding *Concord* are unavailing. OpenAI asserts that nothing suggests Judge Van Keulen would have ordered production of the entire 5-million log sample if Anthropic had raised privacy concerns. To that point, nothing suggests that Judge Van Keulen would not have directed the same outcome either. And to privacy concerns specific to the differences between the types and number of logs, those differences are accounted for by other protections that are in place in this case, as discussed below.

[8] As OpenAI argued in October:

> These outputs are also critically relevant to the fourth fair use factor and damages because the Times has alleged that infringing outputs cause it great harm. The Times has also signaled that it may pursue a free-floating market harm theory not tied to specific infringing outputs, but rather to broader allegations that "pink slime organizations" harm The Times by "flooding the market with low-quality content." This theory is inconsistent with hornbook copyright law. But, to the extent The Times asserts this theory, it must produce the specific

OpenAI's substantial non-infringing uses defense. *See In re OpenAI, Inc., Copyright Infringement Litig.*, --- F. Supp. 3d ---, 2025 WL 2691297, at *5 (S.D.N.Y. Sept. 19, 2025) ("What is relevant is how <u>other consumers</u> are using Defendants' LLMs, and whether <u>those consumers</u> are using Defendants' LLMs for substantially non-infringing purposes or not.") (emphasis in original). (*See also* ECF 164 at 36). Thus, the 20 Million ChatGPT Logs are clearly relevant to News Plaintiffs' output claims to the extent that they contain partial or whole reproductions of News Plaintiffs' copyrighted works, and to OpenAI's affirmative defenses to the extent that they contain other user activity—and News Plaintiffs are entitled to discovery on both.

Production of the 20 Million ChatGPT Logs is also proportional to the needs of the case. The total universe of retained consumer output logs is in the tens of billions. The 20 million sample here represents less than 0.05% of the total logs that OpenAI has retained in the ordinary course of business. (*See* ECF 22-1 at 9). Moreover, the burden of production is minimal at this point; OpenAI has completed (or nearly completed) the extensive process of applying its internal de-identification tool—which OpenAI itself has lauded as significantly more effective at removing both personally identifiable information <u>and</u> private information, (*see* ECF 435 at 1)— and OpenAI has articulated no additional burden with producing the de-identified output logs to News Plaintiffs.

The Court recognizes that the privacy considerations of OpenAI's users are sincere. However, such considerations are only one factor in the proportionality analysis, and cannot predominate where there is clear relevance and minimal burden. *See* The Sedona Conference,

---

"pink slime" outputs at issue so that OpenAI can test The Times's assertions in discovery.

(ECF 635 at 3) (internal citations omitted).

Primer on Social Media, 20 SEDONA CONF. J. 1, 27-28 (2019) (explaining that privacy

considerations are just one factor in assessing the proportionality of discovery). Moreover,

there are multiple layers of protection in this case precisely because of the highly sensitive and

private nature of much of the discovery that is exchanging hands. The first is reducing the total

output logs subject to discovery from tens of billions to 20 million. Second, OpenAI has de-

identified the 20 Million ChatGPT Logs (or will soon complete de-identification). Third,

consumers' privacy is safeguarded by the existing protective order in this case, and by

designating the output logs as "attorneys' eyes only." As I stated in the November 7 order,

OpenAI has "failed to explain how its consumers' privacy rights are not adequately protected by

… the existing protective order in this multidistrict litigation" and OpenAI's de-identification

process. (ECF 734). OpenAI failed to do so again in its motion for reconsideration, (*see* ECF 742),

and asserts only in passing in its reply (and without support) that the Protective Order is "not a

reason to disclose" the 20 Million ChatGPT Logs. (ECF 752). Fourth, the November 7 order

directed the parties to continue meeting and conferring on additional ways to protect OpenAI's

consumers' privacy. (ECF 734 at 2).

Nor does *Nichols v. Noom*—which OpenAI cited in its supplemental briefing and the

Court already considered—suggest that such privacy protections are insufficient here. 20-CV-

3677 (LGS) (KHP), 2021 WL 1997542 (S.D.N.Y. May 18, 2021). There, Judge Parker held that it

was proper to apply search terms to the 2,500-log sample and to only produce logs that

returned as hits, as the remaining logs did not contain relevant information. *Id.* at *3. As

explained above, however, the logs at issue here are relevant for multiple purposes, including

OpenAI's defenses. Moreover, Judge Parker explicitly stated that redaction of personally

identifying information and medical information was not necessary given the protective order

in that case, and given the ability of Noom to designate the chats attorney's eyes only. *Id.* at 4. Here, on top of the Protective Order and OpenAI's ability to designate the logs as attorney's eyes only, OpenAI will complete its de-identification of the 20 Million ChatGPT Logs before production to News Plaintiffs. It would seem then that the privacy protections here are greater than those in *Noom*.[9]

Thus, given that the 20 Million ChatGPT Logs are relevant and that the multiple layers of protections will reasonably mitigate associated privacy concerns, production of the entire 20 million log sample is proportional to the needs of the case. Accordingly, OpenAI's motion for reconsideration is **DENIED.**

IV.   **CONCLUSION**

For the foregoing reasons, OpenAI's motion for reconsideration is **DENIED.** OpenAI is directed to produce the 20 Million ChatGPT Logs within 7 days of completing the de-identification process. Again, the parties are directed to continue meeting and conferring regarding any additional measures that the parties can take to effectively protect the privacy interests of OpenAI's consumers without further delaying discovery in this case.

SO ORDERED.

s/ Ona T. Wang

Dated: December 2, 2025   **Ona T. Wang**
     New York, New York   United States Magistrate Judge

---

[9] OpenAI's argument is also belied by its de-identification of the entire 20 million log sample. If OpenAI never intended to produce the entire 20 Million ChatGPT Logs, it would not have (and should not have) spent the time and money de-identifying the entire 20 million log sample. Either OpenAI initially intended to produce the 20 million logs to News Plaintiffs and changed its mind, for one reason or another; or OpenAI never intended to produce the logs and de-identified the entire 20 million either as a discovery tactic, or for some other reason that has not been identified. Neither bode well for OpenAI.