November 14, 2025

VIA ECF
Hon. Ona T. Wang
U.S. Magistrate Judge for the Southern District of New York
Daniel Patrick Moynihan United States Courthouse
500 Pearl Street
New York, NY 10007

Re: *In re: OpenAI, Inc., Copyright Infringement Litigation*, No. 1:25-md-3143,
This document relates to: No. 1:23-cv-11195

Dear Judge Wang:

News Plaintiffs continue to misrepresent the facts with contradictory and misleading assertions. This brief corrects the record. OpenAI's request for reconsideration has nothing to do with "evad[ing] production of evidence." *Contra* ECF 746 at 1. To the contrary, News Plaintiffs are the ones evading the process: OpenAI weeks ago advanced a proposal to identify and make available potentially relevant logs via n-gram matching and other searches, ECF 717-1, and News Plaintiffs have refused to explain why that proposal is insufficient or even to consider it at all. OpenAI seeks reconsideration because The Times' demand for wholesale production would unnecessarily violate the privacy interests of millions of ChatGPT users worldwide.

First, News Plaintiffs have repeatedly mischaracterized the de-identification process, including by claiming it is "anonymization" and suggesting that it is sufficient to eliminate all privacy concerns. ECF 656 at 2; ECF 719 at 1; ECF 746 at 1; *see also* ECF 734 at 1 (Court referring to OpenAI's de-identification process as "exhaustive"). That is false: as OpenAI has now explained in ***two*** sworn witness declarations (ECF 435-1 ¶8, ECF 683 ¶3), the de-identification that has been proposed here is not designed to scrub information that is private but not directly identifying—like confidential source code, proprietary trading strategies, medical details, or financial information. This is exactly the type of highly private information The Times cited when resisting production of its own logs. ECF 475 at 4–5. OpenAI has made this point at least three times. ECF 679 at 2; ECF 717 at 4–5; ECF 742 at 2. News Plaintiffs have never responded to it.

Second, News Plaintiffs notably make no attempt to defend their request on relevance or proportionality grounds. Nor could they: it would make no sense for News Plaintiffs to argue that production of ***20 million*** user logs belonging to ***third parties*** is appropriate under Rule 26, when they have so recently argued that production of only ***82,150*** user logs generated by ***a party to this lawsuit*** is disproportionate and overly burdensome. ECF 475 at 4–5.

Third, News Plaintiffs try to back away from their prior concession that at least 99.99% of the 20 million ChatGPT records are irrelevant to their claims. ECF 746 at 2. But the record is clear: in July, News Plaintiffs and their expert insisted that the "prevalence" of outputs containing relevant information within the sample is likely between 0.001% and 0.006%. ECF 394 at 2–3; ECF 394-1 ¶4. It follows that, by News Plaintiffs' own admission, ***at least 99.996% to 99.999%*** of ChatGPT conversations likely do ***not*** contain information relevant to their claims.

Fourth, News Plaintiffs now claim that "14-24%" of ChatGPT logs are relevant to their claims based on a paper finding that between 14% and 24% of ChatGPT conversations were classified into a category titled "Seeking Information."[1] Wrong. The fact that someone used ChatGPT to "seek[] information"—about a health condition, the number of stars in the Milky Way, or how to poach an egg—does not mean that conversation log is relevant to News Plaintiffs' claims. At best, this is yet another concession that the vast majority (up to 86%) of the logs are irrelevant.

Fifth, News Plaintiffs claim that the Protective Order is sufficient to eliminate any privacy risk. But the existence of a protective order is not a reason to disclose a trove of irrelevant private information so that it can be pored over by plaintiffs' lawyers and consultants. No court would require Google to hand over millions of irrelevant Gmail messages, even under a protective order. There is no reason why a different standard should apply to ChatGPT conversation logs.

Sixth, News Plaintiffs now abandon their argument that Judge Van Keulen's orders in *Concord* support wholesale production without narrowing for relevance—*i.e.*, the sole basis for the Court's order. ECF 734 at 2. They do not dispute that Anthropic in *Concord* affirmatively offered to produce 5 million prompt-output pairs ***without*** "filter[ing] for relevance." *Concord* ECF 341 at 10.[2] Nor do they dispute that Judge Van Keulen never had occasion to rule on whether wholesale production (as opposed to narrowing for relevance) was appropriate under Rule 26.[3]

Finally, News Plaintiffs suggest that this Court has ruled on this question "repeatedly." ECF 746 at 1–2. Again, wrong. The parties' briefing to date has focused primarily on the issue of ***preservation***. ECF 717 at 2–3 (explaining this distinction). This is a new issue that News Plaintiffs first raised by motion on October 15, 2025. *See* ECF 656. None of the Court's prior orders address this issue, or (for that matter) any issue relating to conversation ***sampling***.

OpenAI again respectfully requests that the Court vacate the order docketed at ECF 734 and instruct the parties to confer regarding OpenAI's proposal for narrowing the logs for relevance. Further, notwithstanding that this Court's order contemplates production "within 7 days of completing [] de-identification," ECF 746 at 3, News Plaintiffs demand that OpenAI produce the data "on November 14, 2025," ECF 746 at 3. That is inappropriate and impossible. De-identification is technically complicated (especially for 20 million records, ECF 435-1 ¶¶8–12) and will not be complete until the end of this month, ECF 742 at 3 n.5. Forcing production before the logs are fully de-identified will drastically exacerbate the privacy risk.[4]

---

[1] *See* ECF 746 at 2; ECF 719 at 5; *see also* Chatterji, et al., *How People Use ChatGPT* (Sept. 15, 2025).

[2] The fact that OpenAI cited *Concord* in the context of a different dispute—about the ***size*** of the sample, not whether it should be ***produced*** wholesale—is entirely inapposite. *Contra* ECF 746 at 2. News Plaintiffs do not dispute that they did not cite *Concord* in support of their demand for wholesale production until their October 30 brief.

[3] Contrary to News Plaintiffs' suggestion, the fact that Latham & Watkins represented Anthropic in *Concord* is a matter of public record. And Latham is not the only law firm in this MDL with other clients in this area. *See, e.g.*, Compl. ¶69 *in Encyclopaedia Britannica, Inc. v. Perplexity AI, Inc.*, No. 1:25-cv-07546, ECF 1 (S.D.N.Y. Sept. 10, 2025) (complaint filed by Davida Brook praising OpenAI as "an example of a leading AI company" that "follows best practices" and "demonstrates how well-intentioned crawlers acting in good faith behave").

[4] Wholesale production of this quantity raises additional privacy concerns not contemplated by OpenAI's original proposal, which assumed that the parties would narrow the logs to a relevant subset before production. If wholesale production is required, OpenAI will use best efforts to address these concerns by augmenting its de-identification process. OpenAI is exploring methods to accomplish this and estimates that it will take an additional two weeks.

Respectfully submitted,

KEKER, VAN NEST & PETERS LLP	LATHAM & WATKINS LLP	MORRISON & FOERSTER LLP
/s/ Edward A. Bayley	*/s/ Margaret Graham*	*/s/ Rose S. Lee*
Edward A. Bayley	*Margaret Graham**	*Rose S. Lee**

cc: All Counsel of Record (via ECF)

*All parties whose electronic signatures are included herein have consented to the filing of this document.