November 12, 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 23-cv-11195

Dear Judge Wang:

On Monday, this Court entered an order requiring OpenAI to hand over to the New York Times and its co-plaintiffs 20 million ChatGPT user conversations, more than 99.99% of which plaintiffs concede have ***nothing to do*** with this case. ECF 734. This data belongs to ChatGPT users all over the world—families, students, teachers, government officials, financial analysts, programmers, lawyers, doctors, therapists, and even journalists—whose private thoughts and confidential business information may now be exposed in this lawsuit. To be clear: anyone in the world who has used ChatGPT in the past three years must now face the possibility that their personal conversations will be handed over to The Times to sift through at will in a speculative fishing expedition.

As OpenAI repeatedly argued before this Court, neither common sense nor the Federal Rules justify the forced production of a massive trove of irrelevant personal user conversations. ECF 679, ECF 717. The Court overruled these important concerns in an order that does not discuss relevance nor proportionality, but nonetheless directs OpenAI to produce millions of conversations belonging to individuals who have no role, voice, or stake in these proceedings.

OpenAI is unaware of any court ordering wholesale production of personal information at this scale. This sets a dangerous precedent: it suggests that anyone who files a lawsuit against an AI company can demand production of tens of millions of conversations without first narrowing for relevance. This is not how discovery works in other cases: courts do not allow plaintiffs suing Google to dig through the private emails of tens of millions of Gmail users irrespective of their relevance. And it is not how discovery should work for generative AI tools either. To the contrary: user privacy is even more essential here.[1] OpenAI is committed to safeguarding that privacy. It cannot stand by while the Times convinces this Court to undervalue the privacy interests of tens of millions of users for the Times's financial benefit.

OpenAI respectfully requests that this Court reconsider its order. The Court's holding assumed that OpenAI had "fail[ed] to explain" why Judge Van Keulen's rulings in *Concord v. Anthropic*

[1] Nils Gilman, *If You Tell ChatGPT Your Secrets, Will They Be Kept Safe?*, New York Times (Nov. 10, 2025) (disclosure of ChatGPT conversations "will destroy the candid relationship that makes A.I. useful for mental health and legal and financial problem-solving").

were not "instructive." ECF 734 at 2. But News Plaintiffs did not even reference *Concord* in their motion. ECF 656. They raised it for the first time in their subsequent Court-ordered simultaneous submission, misleadingly referring to it as "the most relevant case on this issue." ECF 719. So OpenAI never had a chance to explain that unlike OpenAI, Anthropic had affirmatively ***proposed*** wholesale production of 5 million user logs without any apparent concern for the privacy implications. The *Concord* court thus never had to "direct[]" wholesale production. *Contra* ECF 734 at 2. To the contrary, the *Concord* order was about how to effectuate an ***already agreed-upon*** production. *Concord* ECF 407 at 4–5. That is grounds for reconsideration. *Shrader v. CSX Transp., Inc.*, 70 F.3d 255, 257 (2d Cir. 1995).

## I. OpenAI Never Had The Opportunity to Explain Why *Concord* Is Inapposite

OpenAI has already surveyed the history of this dispute. ECF 717 at 3–4. As relevant here: News Plaintiffs' discovery requests are all limited to conversation logs that (e.g.) are "related to Times content." *See, e.g.*, *NYT* ECF 379-8 at 2. The parties have been working to satisfy those requests by sampling conversation logs. Towards the end of that process, News Plaintiffs filed a motion with a new demand: that instead of finding and producing logs that are "related to Times content," OpenAI should hand over the ***entire*** 20 million-log sample "via hard drive." ECF 656 at 3.

OpenAI pushed back on relevance and proportionality grounds, highlighting privacy concerns. ECF 679. On October 29, 2025 the Court ordered the parties to simultaneously file new briefs on the issue. While OpenAI's brief reprised the same arguments from its initial opposition, ECF 717, News Plaintiffs' brief ***for the first time*** cited Judge Van Keulen's August order in *Concord v. Anthropic*, claiming it is "the most relevant case on this issue," ECF 719 at 4, even though it was not mentioned in their initial brief, ECF 656.

On November 10, 2025, the Court granted News Plaintiffs' motion. ECF 734. The order did not address relevance nor proportionality. Nor did it acknowledge News Plaintiffs' concession that at least 99.99% of the conversation logs are irrelevant to their claims and not responsive to any discovery requests. It responded to OpenAI's privacy concerns by citing OpenAI's intent to "de-identif[y]" the at-issue conversations, but did not acknowledge OpenAI's sworn witness declaration explaining that the de-identification process is not intended to remove information that is non-identifying but may nonetheless be private, like a *Washington Post* reporter's hypothetical use of ChatGPT to assist in the preparation of a news article. ECF 683 ¶3; *see also* ECF 475 at 4–5 (NYT claiming such logs are private).[2] Nor did the order acknowledge that other courts have rejected similar requests, holding that plaintiffs "are not entitled to" user logs that "do not contain relevant communications within the meaning of Rule 26." ECF 717 at 5 (discussing *Nichols v. Noom*, No. 20-cv-3677, 2021 WL 1997542, at *3 (S.D.N.Y. May 18, 2021)).

Instead, the order relied on Judge Van Keulen's rulings from the *Concord v. Anthropic* case directing production of a "5-million record sample." ECF 734 at 2. And it faulted OpenAI for

[2] Natasha Singer, *With a Few Bits of Data, Researchers Identify 'Anonymous' People*, N.Y. Times (Jan. 29, 2015) ("Even when real names and other personal information are stripped from big data sets, it is often possible to use just a few pieces of the information to identify a specific person.").

"fail[ing] to explain why [those *Concord* orders]" are not "instructive," even though OpenAI ***never*** had the opportunity to do so.

## II. *Concord* Is Inapposite

In any event, the *Concord* orders are inapposite for two critical reasons. First, Judge Van Keulen never considered the privacy implications at issue here because Anthropic did not raise them. To the contrary, Anthropic from the outset proposed to voluntarily hand over a "statistical sample of prompt and output data" so that plaintiffs' "experts can use [it] to analyze typical [] user behavior." *Concord* ECF 302 at 10; *see also* *Concord* ECF 341 at 10 (Anthropic: "prompts and outputs in the sample ***will not be filtered for relevance***" (emphasis added)). The cited *Concord* order was not about whether wholesale production of the sample was appropriate; it was about the mechanism through which Anthropic would effectuate an ***already agreed-upon*** production. *Concord* ECF 407 at 4–5.[3] Nothing about that order suggests that Judge Van Keulen would have ordered wholesale production had Anthropic raised the privacy concerns that OpenAI has raised throughout this case.

Second, the logs at issue in the *Concord* litigation were nothing like the logs at issue here. The logs in *Concord* were "prompt-output pairs"—*i.e.*, a single user prompt followed by a single model output. *Concord* ECF 377 at 1. The logs at issue here are ***complete conversations***: each log in the 20 million sample represents a complete exchange of multiple prompt-output pairs between a user and ChatGPT. Disclosure of those logs is thus much more likely to expose private information, in the same way that eavesdropping on an entire conversation reveals more private information than a 5-second conversation fragment. Moreover, there are more logs at issue here than in *Concord*: this Court's order could require production of up to 80 million prompt-output pairs[4]: ***16x more*** than the 5 million at issue in *Concord* (and almost ***1000x more*** than the 81,000 pairs this Court found too burdensome to be produced after The Times objected, ECF 604 at 12).

Nothing in the cited *Concord* order supports News Plaintiffs' demand for wholesale production, nor undercuts OpenAI's proposal to narrow based on relevance. On the other hand, the only on-point precedent cited by either party ***rejected*** wholesale production in favor of narrowing. *Nichols v. Noom*, No. 20-cv-3677, 2021 WL 1997542, at *3 (S.D.N.Y. May 18, 2021).

## III. The Court Should Vacate The Order

OpenAI respectfully requests that the Court vacate the order and order News Plaintiffs to respond to OpenAI's proposal for identifying relevant logs. ECF 717 at 5 & Ex. A. In the interim, OpenAI requests the Court stay the order pending resolution of this motion because production will irreversibly harm the privacy interests discussed above.[5]

---

[3] That order rejected the plaintiffs' demand that Anthropic "produce copies of the[] records," *Concord* ECF 399 at 4, and instead adopted a proposal to make the data available via a secure environment, *Concord* ECF 407 at 4.

[4] Conservatively assuming four prompt-output pairs per conversation.

[5] OpenAI anticipates completing the de-identification process before the end of the month and, if required, providing access via a secure environment within 7 days. ECF 734 at 2.

Respectfully submitted,

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