

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

IN RE:

25-md-03143 (SHS) (OTW)

OPENAI, INC.,
COPYRIGHT INFRINGEMENT LITIGATION

Hon. Sidney H. Stein

This Document Relates To:

THE NEW YORK TIMES COMPANY v.
MICROSOFT CORPORATION, et al., No. 23-
cv-11195

**MEMORANDUM OF LAW IN SUPPORT OF OPENAI DEFENDANTS' OBJECTION
TO ORDER COMPELLING PRODUCTION OF CHATGPT CONVERSATION LOGS
AT MDL ECF 734 PURSUANT TO FEDERAL RULE OF CIVIL PROCEDURE 72(a)**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. BACKGROUND	2
A. ChatGPT Conversation Data Discovery Requests and Sampling.....	2
B. Production of ChatGPT Conversations and the Challenged Order	3
III. STANDARD OF REVIEW	5
IV. ARGUMENT	6
A. The Order is clearly erroneous because it compels discovery far beyond the scope of News Plaintiffs’ discovery requests	6
B. The Order fails to apply Rule 26 and the required proportionality analysis.....	6
C. The Order is also disproportionate because the vast majority of conversations are unlikely to be probative but their production will impact the privacy rights of millions of ChatGPT users	7
i. Privacy interests weigh heavily against disclosure of millions of unrelated conversation records	8
ii. The needs of this case center on News Plaintiffs’ infringement allegations and do not warrant the disclosure of conversations unrelated to their content	10
D. The Court should narrow production to all <i>responsive</i> conversations, as identified by search terms tailored to News Plaintiffs’ discovery requests	12
V. CONCLUSION.....	14

TABLE OF AUTHORITIES

Page(s)

CASES

Concord Music Grp. v. Anthropic PBC,
 No. 24-CV-3811 (EKL) (SVK), 2025 WL 1482734 (N.D. Cal. May 23, 2025).....5

Coventry Cap. U.S. LLC v. EEA Life Settlements Inc.,
 333 F.R.D. 60 (S.D.N.Y. 2019)5, 6, 7

Gonzales v. Google, Inc.,
 234 F.R.D. 674 (N.D. Cal. 2006).....10

Matter of New York Times Co.,
 828 F.2d 110 (2d Cir. 1987).....8

Nichols v. Noom,
 2021 WL 1997542 (S.D.N.Y. May 18, 2021)11, 13

R.F.M.A.S., Inc. v. So,
 748 F. Supp. 2d 244 (S.D.N.Y. 2010).....5

S.E.C. v. Rajaratnam,
 622 F.3d 159 (2d Cir. 2010).....8, 9, 11

Treppel v. Biovail Corp.,
 233 F.R.D. 363 (S.D.N.Y. 2006)13

RULES

Fed. R. Civ. P. 26.....6, 7, 8

Fed. R. Civ. P. 26(b)(1).....2, 7

Fed. R. Civ. P. 26(b)(2)(C)(iii)7

Fed. R. Civ. P. 376

Fed. R. Civ. P. 37(a)(3)(B)(iv).....6

Fed. R. Civ. P. 72(a)1, 5, 14

OTHER AUTHORITIES

Hon. Ona T. Wang, *Model Joint Electronic Discovery Submission and Proposed Order* (Mar. 8, 2018) § 6, available at https://nysd.uscourts.gov/sites/default/files/practice_documents/otwModelJointElectronicDiscoverySubmissionAndProposedOrder.pdf13

Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701 (2010).....10

The Sedona Principles, Third Edition, 19 Sedona Conf. J. 1, Principle 11, 164-68 (2018), available at https://www.thesedonaconference.org/sites/default/files/3-2_Sedona_Principles_3d_Ed.pdf.....13

I. INTRODUCTION

OpenAI respectfully objects to Magistrate Judge Wang’s Order¹ (the “Order”) directing OpenAI to produce to the New York Times and its co-plaintiffs² (collectively, “News Plaintiffs”) 20 million private ChatGPT user conversations, more than 99.99% of which News Plaintiffs concede do not relate to their claims that ChatGPT outputs infringe their content. ECF 734. The Order invades the privacy rights of millions of non-party ChatGPT users, compelling the disclosure of their private conversations without any showing of relevance to News Plaintiffs’ claim of infringement. Such an overbroad intrusion would set a troubling precedent, at odds with privacy principles that federal courts have long recognized and protected. Indeed, unless reversed, the Order will sanction one of the largest transfers of private personal information ever by a United States Federal District Court—and, as explained below, for no legitimate purpose. The Order is clearly erroneous, contrary to law, and should be set aside for the parties to develop a more narrowly tailored approach to satisfying the News Plaintiffs’ discovery interests while balancing the privacy concerns of millions of users worldwide.³

This Order arises from News Plaintiffs’ efforts to seek discovery to support their allegation that ChatGPT is capable of “generat[ing] output that recites [News Plaintiffs’] content verbatim,” *NYT* ECF 170 ¶4 (citing Exhibit J). But News Plaintiffs concede that only a small fraction (between 0.001% and 0.006%) of the 20 million conversations potentially support that hypothesis. As a result, the overbroad scope of the Order and its impact on user privacy are disproportionate

¹ OpenAI has moved for reconsideration of the Order. ECF 742. Under FRCP 72, however, OpenAI cannot wait for that ruling and must file this objection to the Order within 14 days of its issuance. FRCP 72(a) (“A party may serve and file objections to the order within 14 days after being served with a copy.”).

² The *Daily News* Plaintiffs, the Center for Investigative Reporting, The Intercept, and the *Ziff Davis* Plaintiffs.

³ Since the entry of the Order, other plaintiffs in this MDL have also demanded to inspect all 20 million conversations implicated by it, further exacerbating the inherent harm to user privacy that is a collateral consequence of the Order.

and not tailored to the discovery actually sought: evidence that ChatGPT outputs infringe News Plaintiffs’ works. Before the Order was entered, OpenAI proposed using traditional e-discovery techniques (*i.e.*, targeted search terms) to identify the subset of 20 million ChatGPT conversations responsive to News Plaintiffs’ discovery requests, which only seek conversations related to their news articles, but News Plaintiffs refused to consider OpenAI’s proposal or explain why it was not workable. Instead, they insisted that OpenAI produce all 20 million conversations without any narrowing for relevance, responsiveness, or proportionality. The Order assented to News Plaintiffs’ demand without acknowledging that 99.99% of the conversations ordered produced are not “relevant to [News Plaintiffs’] claim” that ChatGPT outputs infringe, or that forcing production of millions of private third-party conversations for the sake of convenience and expediency—notwithstanding the considerable impact to user privacy—is not “proportional to the needs of the case.” FRCP 26(b)(1).

OpenAI therefore respectfully requests that the Court vacate the Order, so that the parties can quickly identify and produce the responsive information that is actually relevant and proportional to the needs of the case as the Federal Rules of Civil Procedure require.

II. BACKGROUND

A. ChatGPT Conversation Data Discovery Requests and Sampling

In May 2024, News Plaintiffs requested inspection and production of “[q]uery, session, and chat logs *related to [their] Content*,” including “user queries paired with responses to those queries.” Bayley Decl. Ex. A, at 2 (emphasis added); Bayley Decl. Ex. B, at 13; *see also* Bayley Decl. Ex. C at 2; Bayley Decl. Ex. D, at 17. OpenAI responded that the data was not readily searchable given the format in which it was stored,⁴ and so the parties would need to derive a

⁴ Due to the significant volume of data generated by the hundreds of millions of ChatGPT conversations each month, OpenAI compresses conversation log data before placing it into long-term, “offline” storage. *See* ECF 43-18. After a

representative sample that could be searched. *See* Declaration of E. Bayley Ex. E. The parties continued to meet-and-confer thereafter for the next several months. *See* Bayley Decl. Exs. F–K (listing correspondence).

On May 20, 2025, News Plaintiffs proposed a sampling methodology contemplating a sample of more than 1.4 billion ChatGPT conversation logs. *See* Bayley Decl. Ex. L, at 3–4. Pursuant to meet-and-confers, in June 2025, OpenAI offered as a compromise to create a sample of up to 20 million conversations from which responsive content could be searched. Bayley Decl. Ex. M. Although News Plaintiffs initially filed a motion seeking to compel OpenAI to retrieve a sample of 120 million conversations, *see* ECF 394, they eventually agreed to proceed with 20 million. ECF 719-2.

In August, News Plaintiffs furnished OpenAI with a list of nearly 20 million conversations, selected from a metadata list OpenAI turned over months earlier, to use as the sample for the searching of records related to News Plaintiffs’ content. *Id.* OpenAI immediately began the extensive and complex process of retrieving, decompressing, processing, and de-identifying those 20 million conversations. *See* ECF 435-1 ¶3 (describing the decompression and deidentification process).

B. Production of ChatGPT Conversations and the Challenged Order

On October 14, 2025, News Plaintiffs demanded for the first time that OpenAI “commit to producing output log data” in its entirety. Bayley Decl. Ex. R, at 9. Before that date, the parties

conversation log is compressed, only a limited amount of metadata about the conversation, such as the date it was created, remains visible and in a searchable format. *See* Bayley Decl. Ex. K, at 2. To view and search the full content of the logs, the logs must first be decompressed. *Id.* In other words, OpenAI cannot, for example, search for every conversation in which a user mentioned “The New York Times” without first decompressing and processing every single one of the billions of ChatGPT conversations in offline storage. Because of these technical challenges, the parties have agreed that the appropriate way to test News Plaintiffs’ output-based infringement claims is to decompress a sample of ChatGPT conversations so that statistically significant conclusions can be drawn about the full set of historical conversation data. *See, e.g.*, Bayley Decl. Ex. H.

had not discussed which conversations within the sample would be made available in response to News Plaintiffs' discovery requests, or the logistics of making these conversations available. ECF 656 at 1. OpenAI explained that, while it could not turn over 20 million conversations wholesale, given user privacy concerns and the fact that the overwhelming majority are not related to News Plaintiffs' content, it would work with News Plaintiffs "to identify the [conversations] that are actually relevant to [their] claims" of output-based copyright infringement and make them available. Bayley Decl. Ex. R, at 8.

The next day, without responding to OpenAI's offer to meet and confer, News Plaintiffs filed a motion seeking to compel OpenAI to produce the entire 20 million sample. ECF 656 at 1. News Plaintiffs falsely asserted that OpenAI "renege[d]" on an agreement to produce the entire sample, even though the parties had yet to discuss these specific issues, much less come to an agreement. News Plaintiffs have not, either in their initial motion or in subsequent briefing, explained why OpenAI's proposal was not sufficient to find the logs responsive to their requests.

At the October 29, 2025 discovery conference, Judge Wang instructed the parties to file supplemental briefs explaining how the present dispute differed from earlier disputes related to data preservation and conversation sampling methodology. Bayley Decl. Ex Q, at 15:24–16:4. OpenAI proposed methods to News Plaintiffs to search for conversations related to their works and offered to discuss the proposal with News Plaintiffs in an attempt to reach a resolution. Bayley Decl. Ex. R, at 1–2. News Plaintiffs refused. *Id.* OpenAI then submitted a supplemental brief detailing the history of the dispute, *see* ECF 717 at 2–4, and reminding the Court of the serious privacy concerns at stake. *See id.* at 4–5 (emphasizing the privacy implications of producing de-identified conversations).

But on November 10, 2025, the Court issued the Order, which directed OpenAI “to produce the 20 million de-identified Consumer ChatGPT Logs[.]” ECF 734 at 2. The Court faulted OpenAI for failing to explain “how its consumers’ privacy rights are not adequately protected” or why the district court’s order in *Concord Music Grp., Inc. v. Anthropic PBC* requiring production of an entire 5-million record sample “is not similarly instructive here.” *Id.* at 1–2; *see also Concord Music Grp. v. Anthropic PBC*, No. 24-CV-3811 (EKL) (SVK), 2025 WL 1482734, at *3–4 (N.D. Cal. May 23, 2025).

OpenAI promptly moved for reconsideration, explaining that it had not addressed the *Concord* order because News Plaintiffs first raised *Concord* in their simultaneous supplemental briefing—not in their original motion to compel at ECF 656—and therefore OpenAI had no opportunity to respond. ECF 742 at 2. OpenAI further explained that the order in *Concord* was inapposite because defendant-Anthropic *voluntarily* offered to produce the entire sample of queries and responses negotiated by the parties in that litigation. *Id.* at 3. As a result, the court in *Concord* did not consider the privacy considerations implicated by compelling production of millions of entire conversations as is the case here. *Id.* OpenAI’s motion for reconsideration remains pending before Judge Wang.

III. STANDARD OF REVIEW

A district court may “modify or set aside” any part of a magistrate judge’s non-dispositive order “that is clearly erroneous or is contrary to law.” Fed. R. Civ. P. 72(a). An order is “contrary to law” if it “fails to apply or misapplies relevant statutes, case law, or rules of procedure.” *Coventry Cap. U.S. LLC v. EEA Life Settlements Inc.*, 333 F.R.D. 60, 64 (S.D.N.Y. 2019) (quoting *R.F.M.A.S., Inc. v. So*, 748 F. Supp. 2d 244, 248 (S.D.N.Y. 2010)). An order is “clearly erroneous” if it provides the reviewing court a “definite and firm conviction that a mistake has been committed.” *Id.*

IV. ARGUMENT

A. The Order is clearly erroneous because it compels discovery far beyond the scope of News Plaintiffs' discovery requests.

The scope for a motion to compel production turns on what was “requested under Rule 34.” Fed. R. Civ. P. 37(a)(3)(B)(iv). Here, News Plaintiffs’ operative discovery requests only asked OpenAI for “chat logs reflecting or analyzing user sessions *related to [News Plaintiffs’] Content[.]*” See, e.g., Bayley Decl. Ex. B at 13 (emphasis added); see also ECF 719 at 2 (News Plaintiffs’ admission that their discovery requests only asked for “output logs that contain or use News Plaintiffs’ copyrighted works”). OpenAI has repeatedly expressed its willingness to provide that discovery from the sample. Indeed, OpenAI has made multiple proposals to apply standard ESI search techniques to identify the 0.001% to 0.006% of records *that are even potentially responsive* to News Plaintiffs’ discovery requests without unnecessarily including millions and millions of private, *unrelated* conversations. See, e.g., ECF 681 at 3; ECF 717 at 5.

The discovery compelled by the Order ignores those offers and imposes a production scope that far exceeds News Plaintiffs’ requests and OpenAI’s obligations under Rule 26. The Order requires production of many millions of private ChatGPT conversations that are *unrelated* to News Plaintiffs’ content, without limitation or even discussion of responsiveness, relevance, or proportionality. See, e.g., ECF 394 at 2–3 (News Plaintiffs estimating that between 0.001% and 0.006% of ChatGPT conversations are likely to be responsive to their discovery requests).

By compelling a production under Rule 37 that far exceeds what was “requested under Rule 34,” the Order is clearly erroneous and should be set aside on this basis alone.

B. The Order fails to apply Rule 26 and the required proportionality analysis.

The Order should also be set aside as “contrary to law” because it “fails to apply . . . rules of procedure” requiring proportionality determinations under Federal Rule of Civil Procedure 26.

Coventry, 333 F.R.D. at 64. Federal courts must consider relevance and proportionality in determining the scope of permissible discovery, as well as limit the frequency or extent of discovery that exceeds that scope. Fed. R. Civ. P. 26(b)(2)(C)(iii) (“On motion *or on its own*, the court *must* limit the frequency or extent of discovery otherwise allowed by these rules or by local rule if it determines that: . . . the proposed discovery is outside the scope permitted by Rule 26(b)(1).”) (emphasis added). Indeed, the Advisory Committee on Civil Rules has stated that “parties *and the court* have a *collective* responsibility to consider the proportionality of all discovery and *consider it in resolving discovery disputes.*” Fed. R. Civ. P. 26(b)(1) advisory committee’s note to 2015 amendment (emphasis added).

The challenged Order offers no discussion, no findings, and no mention of proportionality before directing OpenAI to produce 20 million private third-party ChatGPT conversations “in whole.” Order at 1. Given the News Plaintiffs’ estimation that less than 0.006% of ChatGPT conversations are even potentially responsive to their discovery requests, the production of *all* chats in the sample is plainly disproportionate. Nor did the Court indicate at the October 29, 2025 discovery status conference that it was considering, or had considered, the proportionality of producing 20 million records under any applicable factor. *See* Bayley Decl. Ex Q, at 16:5–17:21.

Because the Order fails to apply the requirements of Rule 26 (e.g., due consideration of relevance and proportionality), it is contrary to law and must be reversed.

C. **The Order is also disproportionate because the vast majority of conversations are unlikely to be probative but their production will impact the privacy rights of millions of ChatGPT users.**

By compelling production of several million private ChatGPT conversations of third parties “in whole”—without any regard to whether they are relevant or responsive to News

Plaintiffs’ own discovery requests, and at the expense of users’ privacy—the Order is clearly disproportionate to the needs of this copyright infringement action under Rule 26.

i. Privacy interests weigh heavily against disclosure of millions of unrelated conversation records.

Simply put, the privacy interests of millions of Americans are at stake in this discovery dispute. Federal courts recognize a general right to privacy that can be raised in response to discovery requests. *See, e.g., S.E.C. v. Rajaratnam*, 622 F.3d 159, 184 (2d Cir. 2010) (“The privacy interests in the instant case merit particular attention given that the disclosure order implicated thousands of conversations of hundreds of innocent parties, and that the district court ordered disclosure . . . without limiting the disclosure to relevant conversations.”); *Matter of New York Times Co.*, 828 F.2d 110, 116 (2d Cir. 1987) (“The job of protecting such [third party] interests rests heavily upon the shoulders of the trial judge, since all the parties who may be harmed by disclosure are typically not before the court.”).

Every day, hundreds of millions of individuals around the world turn to ChatGPT for highly personal or sensitive purposes, such as drafting an intimate letter to a partner, requesting information on one’s immigration status, managing personal finances, seeking to develop a proprietary trading strategy, or conducting investigative journalism. *See, e.g.,* ECF 475 at 5 (the Times asserting reporters’ privilege over ChatExplorer logs); *see also, e.g.,* Bayley Decl. Ex. U; Bayley Decl. Ex. V (opining that the disclosure of ChatGPT conversations “will destroy the candid relationship that makes A.I. useful for mental health and legal and financial problem-solving”). As a result of the Order, however, anyone in the world who has ever used ChatGPT in the past several years must now face the possibility that these personal conversations will be handed over “in whole” to News Plaintiffs. News Plaintiffs’ insistent attempts to take possession of 20 million of these deeply sensitive and private conversations—the vast majority, and potentially all, of which

are untethered to the issues in this copyright dispute—implicate privacy interests on a scale that would hardly be proportionate in any context.

Unfortunately, the Order addresses these legitimate privacy concerns only briefly, citing the existence of the current protective order and OpenAI’s de-identification procedures that are designed to remove personally identifiable information from the conversations. Order, at 1–2; *see also* ECF 683 at ¶3. Although these measures provide the bare minimum protections needed for disclosure of the subset of conversations responsive to News Plaintiffs’ discovery requests, they cannot justify the sweeping production of many millions of *unrelated* conversations from innocent and unaffiliated individuals.

On this point, the Second Circuit’s decision in *S.E.C. v. Rajaratnam* is instructive. 622 F.3d 159 (2d Cir. 2010). There, the defendants appealed a district court discovery order compelling them to disclose thousands of wiretapped conversations. *Id.* at 164. The Second Circuit reversed, concluding that the district court “clearly exceeded its discretion by failing to limit the disclosure of the wiretapped conversations to *relevant* conversations” because the “SEC does not have a right of access to *irrelevant* conversations, even if lawfully intercepted[.]” *Id.* at 187 (emphasis added). As the Second Circuit observed, the “district court’s protective order limiting disclosure to the SEC and other parties to the litigation was not sufficient to protect the privacy interests at stake.” *Id.* at 186 n.27. So too here: the Order’s reliance on a protective order is not sufficient to protect the privacy interests implicated by the production of *several million* private and irrelevant ChatGPT conversations.

Likewise, OpenAI’s de-identification process, while meaningful, does not and cannot absolve News Plaintiffs of all the other privacy intrusions that would result from rummaging through several million users’ personal conversations, even if those conversations are not directly

affiliated with a particular individual. *See Gonzales v. Google, Inc.*, 234 F.R.D. 674, 687 (N.D. Cal. 2006) (rejecting argument that “there are no privacy issues” absent “identifiable information”). Put simply, de-identification does not eliminate concerns about the sensitivity of data. Indeed, it has been well-documented that de-identification is insufficient on its own to guarantee anonymity, as it may still be possible to trace de-identified information back to the individuals. *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701 (2010); *see also* Bayley Decl. Exs S–T. For instance, in August 2006, after America Online (AOL) publicly released the search queries for 650,000 anonymized users of AOL’s search engine, two of the Times’ own reporters were able to deduce the identity of one of these individuals based on her de-identified search queries. Bayley Decl. Ex. T. As these reports illustrate, although de-identification is an important and necessary tool, it cannot fully address the substantial privacy concerns associated with producing millions of personal conversations of individuals who are not parties to this case. *See* ECF 683 at ¶ 3.

ii. The needs of this case center on News Plaintiffs’ infringement allegations and do not warrant the disclosure of conversations unrelated to their content.

As part of the proportionality analysis, courts must weigh the privacy intrusion of production against the needs of the case. The News Plaintiffs’ discovery requests seek conversations that specifically concern their copyrighted content, and News Plaintiffs have never explained why they need additional conversations, unrelated to their content. Users’ private ChatGPT conversations, therefore, must first be identified as related to these requests and proportional to the needs of this case before being produced, which is precisely why OpenAI planned to use search terms, which News Plaintiffs have not shown are insufficient.

District courts routinely account for these privacy interests when considering the production of highly personal technological interactions and have ordered preliminary searches, before production, to ensure only necessary private records are produced. For instance, in *Nichols v. Noom*, involving facts very similar to those in this case, Judge Parker voiced “concern[s] about the potentially large amount of highly personal, irrelevant information that would be produced” 2021 WL 1997542, at *3 (S.D.N.Y. May 18, 2021). There, as here, the plaintiffs had sought to compel an unfiltered production of chat logs from Noom’s automated weight-loss coach, including a “potentially large amount of highly personal, irrelevant information” about users’ “weight-loss journey, food anxieties, life stressors,” and so on. *Id.* at *1-3. Rejecting that proposal in favor of using preliminary search terms to identify relevant records, Judge Parker held that the plaintiffs “do not actually need to see the chats for the remainder of the 2,500 users (the “non-complaining subset”). . . . *Indeed, they are not entitled to them*, as they do not contain relevant communications within the meaning of Rule 26.” *Id.* at *3 (emphasis added). Judge Parker’s reasoning mirrors the Second Circuit’s conclusions in *Rajaratnam*: a requesting party “does not have a right of access to irrelevant conversations[.]” 622 F.3d at 187.

Such reasoning applies with even greater force here, with **8,000 times more** private conversations at stake. News Plaintiffs are not entitled to (and have never justified any actual need for) millions of private ChatGPT conversations that are unrelated to their content and not responsive to their own discovery requests. For this reason, the Order—which requires production of 20 million private conversations without regard to relevance and proportionality—is clearly erroneous and should be set aside.

D. The Court should narrow production to all responsive conversations, as identified by search terms tailored to News Plaintiffs' discovery requests.

Given the significant third-party privacy interests at stake, the Court should require that parties first identify responsive conversations prior to production so as to avoid the needless intrusion into the private lives of several million ChatGPT users. This baseline proposition, reflected in the express language of News Plaintiffs' own discovery requests, is consistent with the Federal Rules of Civil Procedure and Second Circuit law.

For their part, News Plaintiffs have never explained why all 20 million conversations in the sample would actually be relevant and proportional to the needs of the case. Instead, News Plaintiffs have sought to justify their overbroad demand on the grounds that have nothing to do with the merits of the case, such as expedience and the convenience of their experts. ECF 719 at 1–2. In other words, by demanding millions of admittedly irrelevant (yet deeply personal) ChatGPT conversations, News Plaintiffs have demonstrated that they are willing to sacrifice the privacy rights of millions of ChatGPT users in pursuit of a strategic advantage in the litigation, not discovery they actually need to advance their case. The Court should not countenance such tactics.

Throughout the course of this dispute, OpenAI has maintained that News Plaintiffs will be able to access all ChatGPT conversations related to News Plaintiffs' content in the sample, *i.e.*, precisely what they had asked for in their discovery requests. To that end, OpenAI has offered to perform multiple searches that would capture the universe of conversations responsive to News Plaintiffs' discovery requests. *See, e.g.*, ECF 681 at 3; ECF 717 at 5.

News Plaintiffs have objected to the application of search terms on the basis that negotiation search terms would somehow reveal News Plaintiffs' work product. ECF 719 at 4. News Plaintiffs' professed fears are unsubstantiated and contradicted by the everyday realities of ESI discovery, including ESI discovery in this case—where search terms were discussed and

negotiated, as usual, for other ESI searches. *See The Sedona Principles, Third Edition*, 19 Sedona Conf. J. 1, Principle 11, 164–68 (2018), available at https://www.thesedonaconference.org/sites/default/files/3-2_Sedona_Principles_3d_Ed.pdf (“A responding party may satisfy its good faith obligations to preserve and produce relevant electronically stored information by using technology and processes, such as sampling, searching, or the use of selection criteria.”). Courts in this district have likewise endorsed search strategies to identify responsive documents and have reprimanded plaintiffs’ “recalcitrance” for refusing to agree to a search methodology to define the scope of production. *Treppel v. Biovail Corp.*, 233 F.R.D. 363, 374 (S.D.N.Y. 2006); *see also Nichols*, 2021 WL 1997542, at *3 (ordering that “search terms be applied to identify relevant chats” and holding that “Plaintiffs do not actually need to see the chats for the remainder of the 2,500 users. . . . Indeed, they are not entitled to them, as they do not contain relevant communications within the meaning of Rule 26.”).

Using search terms to identify conversations related to News Plaintiffs’ content would not reveal any more work product than the typical exchange of ESI search terms expressly required by some judges in this district. *See, e.g., Hon. Ona T. Wang, Model Joint Electronic Discovery Submission and Proposed Order* (Mar. 8, 2018) § 6, available at https://nysd.uscourts.gov/sites/default/files/practice_documents/otwModelJointElectronicDiscoverySubmissionAndProposedOrder.pdf (“The parties have discussed methodologies or protocols for the search and review of electronically stored information, as well as the disclosure of techniques to be used. Some of the approaches that may be considered include: the use and exchange of keyword search lists, ‘hit reports,’ and/or responsiveness rates[.]”).

The Court should therefore set aside the Order compelling OpenAI to produce 20 million conversations “in whole,” and instead direct the parties to meet-and-confer on suitable searches to

identify the conversations out of the set of 20 million records that are actually responsive to News Plaintiffs' discovery requests and proportional to the needs of the case while minimizing the privacy burdens on third-parties.

V. CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that the Court set aside the Order under Rule 72(a).

Respectfully submitted,

/s/ Margaret Graham

LATHAM & WATKINS LLP

Andrew M. Gass (*pro hac vice*)

andrew.gass@lw.com

505 Montgomery Street, Suite 2000

San Francisco, CA 94111

Telephone: 415.391.0600

Margaret Graham

margaret.graham@lw.com

Sarang V. Damle

sy.damle@lw.com

Luke A. Budiardjo

luke.budiardjo@lw.com

1271 Avenue of the Americas

New York, NY 10020

Telephone: 212.906.1200

Elana Nightingale Dawson (*pro hac vice*)

elana.nightingaledawson@lw.com

555 Eleventh Street, NW, Suite 1000

Washington, D.C. 20004

Telephone: 202.637.2200

/s/ Rose S. Lee

MORRISON & FOERSTER LLP

Joseph C. Gratz (*pro hac vice*)

jgratz@mofo.com

Tiffany Cheung (*pro hac vice*)

tcheung@mofo.com

Caitlin Sinclair Blythe (*pro hac vice*)

cblythe@mofo.com

425 Market Street

San Francisco, CA 94105

Telephone: 415.268.7000

Rose S. Lee (*pro hac vice*)*

roselee@mofo.com

707 Wilshire Boulevard, Suite 6000

Los Angeles, CA 90017

Telephone: 213.892.5200

/s/ Edward A. Bayley

KEKER, VAN NEST & PETERS LLP

Robert A. Van Nest (*pro hac vice*)

rvannest@keker.com

Edward A. Bayley (*pro hac vice*)*

ebayley@keker.com

Paven Malhotra

pmalhotra@keker.com

Michelle S. Ybarra (*pro hac vice*)

mybarra@keker.com

633 Battery St.

San Francisco, CA 94111

Telephone: 415.391.5400

Attorneys for OpenAI Defendants

* All parties whose electronic signatures are included herein have consented to the filing of this document, as contemplated by Rule 8.5(b) of the Court's ECF Rules and Instructions.

CERTIFICATE OF COMPLIANCE

In accordance with Local Civil Rule 7.1(c), I certify that the foregoing Memorandum of Law is 4,212 words, exclusive of the caption page, table of contents, table of authorities, and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this memorandum.

Dated: November 24, 2025

/s/ Margaret Graham

Margaret Graham