

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

IN RE:

OPENAI, INC.,
COPYRIGHT INFRINGEMENT LITIGATION

This Document Relates To:

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

25-md-3143 (SHS) (OTW)
Hon. Sidney H. Stein

**MEMORANDUM OF LAW IN SUPPORT OF
OPENAI DEFENDANTS' OBJECTION TO PRESERVATION ORDER
AT MDL ECF 33 PURSUANT TO FEDERAL RULE OF CIVIL PROCEDURE 72(a)**

TABLE OF CONTENTS

| | Page |
|---|-------------|
| I. INTRODUCTION | 1 |
| II. BACKGROUND | 4 |
| A. OpenAI and Its Services | 4 |
| B. News Plaintiffs’ Suits..... | 7 |
| C. Output Log Discovery and Preservation Dispute | 8 |
| D. Motion Practice and Preservation Order..... | 9 |
| III. STANDARD OF REVIEW | 11 |
| IV. ARGUMENT..... | 11 |
| A. Ongoing Preservation Will Not Serve A Useful Purpose..... | 12 |
| B. The Order Is Not Proportional to the Needs of the Case | 15 |
| C. The Order Rests on False and Disproven Factual Assumptions | 16 |
| V. CONCLUSION..... | 18 |

TABLE OF AUTHORITIES

Page(s)

CASES

Al Otro Lado, Inc. v. Nielsen,
328 F.R.D. 408 (S.D. Cal. 2018)3, 12, 15, 16

Boudreau v. Petit,
No. 17-cv-301, 2025 WL 553975 (D.R.I. Feb. 19, 2025)4, 15

Coventry Capital U.S. LLC v. EEA Life Settlements Inc.,
333 F.R.D. 60 (S.D.N.Y. 2019)11

Gill v. Gulfstream Park Racing Ass’n., Inc.,
399 F.3d 391 (1st Cir. 2005).....16

Gonzales v. Google, Inc.,
234 F.R.D. 674 (N.D. Cal. 2006).....16

Griffith v. TikTok, Inc.,
No. 5:23-cv-00964, 2024 WL 5279224 (C.D. Cal. Dec. 24, 2024).....14

In re Application of Chevron Corp.,
736 F. Supp. 2d 773 (S.D.N.Y. 2010) (Kaplan, J.)..... *passim*

Riley v. California,
573 U.S. 373 (2014).....15

RULES

Fed. R. Civ. P.
37.....3, 15
37(e)11
72(a)11

OTHER AUTHORITIES

Amanda Caswell, *7 ChatGPT Prompts I'm Using to Improve My Gift Giving - Here's How to Try Them*, Tom's Guide (Dec. 3, 2024),
<https://www.tomsguide.com/ai/chatgpt/7-chatgpt-prompts-im-using-to-improve-my-gift-giving-heres-how-to-try-them>.....3

Karen Bennett, *How to Make a Budget Using ChatGPT*, Bankrate (Feb. 10, 2025), <https://www.bankrate.com/personal-finance/how-to-make-budget-with-chatgpt/#list-out-expenses>.....3

| | Page(s) |
|---|----------------|
| The New York Times Editorial Board, <i>Privacy Cannot Be a Casualty of the Coronavirus</i> , N.Y. Times (Apr. 7, 2020), https://www.nytimes.com/2020/04/07/opinion/digital-privacy-coronavirus.html | 1 |
| OpenAI, <i>API Platform</i> , https://openai.com/api/ | 5, 6 |
| OpenAI, <i>API Reference</i> , https://platform.openai.com/docs/api-reference/introduction | 5 |
| OpenAI, <i>ChatGPT</i> , https://openai.com/chatgpt/overview/ | 5 |
| OpenAI, <i>Data Controls in the OpenAI Platform</i> , https://platform.openai.com/docs/guides/your-data | 6 |
| OpenAI, <i>Endex Builds the Future of Financial Analysis, Powered by OpenAI’s Reasoning Models</i> , https://openai.com/index/endex/ | 6 |
| OpenAI, <i>Enterprise Privacy at OpenAI</i> (Oct. 31, 2024), https://openai.com/enterprise-privacy | 7 |
| OpenAI, <i>Harvey</i> , https://openai.com/index/harvey/ | 6 |
| OpenAI, <i>Healthify</i> , https://openai.com/index/healthify/ | 6 |
| OpenAI, <i>Hebbia’s Deep Research Automates 90% of Finance and Legal Work, Powered by OpenAI</i> , https://openai.com/index/hebbia/ ; | 6 |
| OpenAI, <i>Minnesota’s Enterprise Translation Office Uses ChatGPT to Bridge Language Gaps</i> (Sept. 26, 2024), https://openai.com/index/state-of-minnesota/ | 6 |
| Samantha Kelly, <i>Who Says Romance is Dead? Couples are Using ChatGPT to Write Their Wedding Vows</i> , CNN Business (April 12, 2023), https://www.cnn.com/2023/04/12/tech/chatgpt-wedding-vows/index.html | 3 |

I. INTRODUCTION

This is a Rule 72 objection to an order that threatens the privacy rights of OpenAI’s hundreds of millions of users around the world. On May 13, 2025, in response to a demand by The New York Times and other news plaintiffs, the court issued a sweeping, unprecedented order requiring OpenAI to “preserve and segregate all output log data” without regard to the privacy decisions of its users—and in the absence of any record evidence that such a sweeping order was warranted. *See MDL ECF 33* (the “Preservation Order”). That order, the scope of which the court clarified at a subsequent hearing, remains largely in place today. As a result, OpenAI is forced to jettison its commitment to allow users to control when and how their ChatGPT conversation data is used, and whether it is retained. Every day the Preservation Order remains in place is another day OpenAI’s users are forced to forgo the privacy protections OpenAI has painstakingly put in place. As The New York Times’s own Editorial Board has said, “Americans . . . should be able to control what happens to their personal data.”¹ The Times’s Editorial Board was right then—and The Times is wrong to demand the opposite now. That is why OpenAI has no choice but to lodge this objection.

The Preservation Order is the result of the demand by News Plaintiffs² that OpenAI “preserve [all] output log data” regardless of its content or the privacy decisions of OpenAI’s users. *See NYT ECF 379*. “While courts almost certainly have inherent authority to order those before them to preserve potentially relevant evidence, that authority ordinarily should not be used in the absence of some substantial basis for believing that it would serve a useful purpose.” *In re*

¹ The New York Times Editorial Board, *Privacy Cannot Be a Casualty of the Coronavirus*, N.Y. Times (Apr. 7, 2020), <https://www.nytimes.com/2020/04/07/opinion/digital-privacy-coronavirus.html>.

² News Plaintiffs refers to The New York Times, Daily News Plaintiffs, and Center for Investigative Reporting.

Application of Chevron Corp., 736 F. Supp. 2d 773, 787 (S.D.N.Y. 2010) (Kaplan, J.). The record is devoid of any non-speculative basis for the Preservation Order, let alone a “substantial” one.

News Plaintiffs’ demand for wholesale preservation was premised on their theory that (a) people allegedly use ChatGPT to obtain copies of News Plaintiffs’ content, and (b) such people are allegedly more likely to delete such conversations in an attempt to “cover their tracks.” The problem with that theory—and the Preservation Order that resulted from it—is that News Plaintiffs have not offered a single piece of evidence supporting it. They have not identified any evidence that anyone (other than the News Plaintiffs themselves) has attempted to obtain their content from ChatGPT. In fact, they have conceded that the opposite is true—that ChatGPT is designed *not* to output their content and that ordinary users do *not* use it to elicit such content. *See* Tr. of Sept. 12, 2024 Proceedings 60:3–11 (admitting that The New York Times’s exhibit containing alleged regurgitations of their articles does not reflect “how your average user is using ChatGPT”); Tr. of May 27, 2025 Proceedings at 9:17–21 (News Plaintiffs conceding that it is hard to elicit News Plaintiffs’ content from ChatGPT). Nor have News Plaintiffs identified a single piece of evidence showing that the data OpenAI’s users choose to delete will contain more conversations containing their content (if any) than the billions upon billions of ChatGPT conversations OpenAI has retained. Simply put, there is no basis in the record supporting News Plaintiffs’ speculation that preservation—especially in this overbroad fashion—is necessary. As a result, there is no basis—let alone a “substantial basis”—in the record for the Preservation Order itself. It should thus be vacated as contrary to law.

The Preservation Order should be set aside for the additional reason that preserving “all output log data” imposes significant burdens on, and corresponding harm to, OpenAI and its users that far outweighs News Plaintiffs’ speculative need for such data. Whether preservation is

warranted turns, in relevant part, on whether requiring preservation is proportionate. *See* Fed. R. Civ. P. 37 advisory committee’s notes to 2015 amendment (“[a] factor in evaluating the reasonableness of preservation efforts is proportionality”); *see also Al Otro Lado, Inc. v. Nielsen*, 328 F.R.D. 408, 416 (S.D. Cal. 2018) (“Preservation is evaluated on a standard of reasonableness, which is in turn informed by proportionality.”). The balance does not, and cannot, favor News Plaintiffs.

Indeed, unlike the unsubstantiated hypothetical underlying News Plaintiffs’ request for the Order, the burdens imposed by the Preservation Order are anything but speculative. Millions use ChatGPT daily to improve their lives and to assist in tasks ranging from the mundane to profoundly personal, entrusting OpenAI with their most private thoughts, concerns, and personal details. Such personal information can include financial data for household budgets,³ intimate aspects of their relationships in wedding vows,⁴ and personal preferences for generating gift ideas.⁵ In the business setting, millions likewise rely on OpenAI’s models, trusting OpenAI with their most confidential, trade secret, and privileged business information. OpenAI has earned global trust through its steadfast commitment to user privacy and data security.

In one swoop, the Preservation Order jettisoned that commitment in favor of News Plaintiffs’ unsupported speculation despite the dearth of any record evidence suggesting that doing so would “serve a useful purpose.” *Chevron*, 736 F. Supp. 2d at 787. It also imposed obligations that, by the Order’s terms, would require months of engineering time and cost [REDACTED] of

³ *See* Karen Bennett, *How to Make a Budget Using ChatGPT*, Bankrate (Feb. 10, 2025), <https://www.bankrate.com/personal-finance/how-to-make-budget-with-chatgpt/#list-out-expenses>.

⁴ Samantha Kelly, *Who Says Romance is Dead? Couples are Using ChatGPT to Write Their Wedding Vows*, CNN Business (April 12, 2023), <https://www.cnn.com/2023/04/12/tech/chatgpt-wedding-vows/index.html>.

⁵ Amanda Caswell, *7 ChatGPT Prompts I’m Using to Improve My Gift Giving — Here’s How to Try Them*, Tom’s Guide (Dec. 3, 2024), <https://www.tomsguide.com/ai/chatgpt/7-chatgpt-prompts-im-using-to-improve-my-gift-giving-heres-how-to-try-them>.

dollars. *MDL ECF 66, Monaco Decl.* ¶¶ 4–5. Requiring preservation of user conversation data in the face of such significant, documented burdens—especially when weighed against the lack of any evidence that preservation is warranted—is the opposite of proportionate, as it fails to balance the burden of preservation against the contents of the documents sought to be preserved. *See Boudreau v. Petit*, No. 17-cv-301, 2025 WL 553975, at *2 (D.R.I. Feb. 19, 2025).

Each day the Preservation Order remains in place is another day that OpenAI is forced to ignore the privacy decisions of its users and to incur significant and burdensome costs—both financial and otherwise. Imposing that kind of burden without any “substantial basis” for “believing that it would serve a useful purpose” is both unjust and contrary to law. *Chevron*, 736 F. Supp. 2d at 787. While OpenAI appreciates the court’s efforts to manage discovery in this complex set of cases, it has no choice but to protect the interests of its users by objecting to the Preservation Order and requesting its immediate vacatur.

II. BACKGROUND

A. OpenAI and Its Services

OpenAI is an AI research and deployment company that develops and releases cutting edge generative AI technology, including ChatGPT and its underlying large language models. Hundreds of millions of people from around the world use ChatGPT and its underlying models daily for everything from creating a business plan to learning about a loved one’s illness to sharing their most private thoughts, feelings, and concerns. Given that array of highly confidential and personal use cases, OpenAI goes to great lengths to protect its users’ data and privacy. It does so first and foremost by honoring its privacy policies and contractual commitments to users. Two of OpenAI’s services, and the corresponding policies and contracts that govern those services, are directly implicated by the Preservation Order: OpenAI’s consumer offering (ChatGPT Free, Plus,

and Pro, referred to hereafter as “ChatGPT Consumer”),⁶ and OpenAI’s application programming interface or “API” Platform.

ChatGPT Consumer is OpenAI’s primary consumer offering. *MDL ECF 66, Monaco Decl.* ¶ 2. It is a software platform accessible (among other places) through the web browser on chatgpt.com for anyone to use, for free and without a login. *Id.* Users who create OpenAI accounts can view their “chat history” of past conversations, similar to the list of conversations in a smartphone messaging application. In part to enable this functionality, OpenAI stores users’ past conversations unless users indicate a contrary preference. *Id.* ¶ 3.

Users can elect to *not* retain particular conversations by manually deleting specific conversations or by starting a “Temporary Chat,” which disappears once closed. *Id.* Users can also delete their OpenAI accounts entirely, including all prior conversation history. *Id.* In these instances, OpenAI follows the user’s direction to purge the account conversation data from its systems within 30 days of deletion. *Id.* Honoring that commitment is a top priority for OpenAI and is absolutely essential to the trust OpenAI’s customers place in its services. *MDL ECF 67, Kechida Decl.* ¶ 3. The reason is simple: users feel more free to use ChatGPT when they know that they are in control of their personal information, including which conversations are retained and which are not.

OpenAI’s API is a business offering that allows developers to access OpenAI’s models for use in their own applications.⁷ Unlike ChatGPT, which has a user-facing interface, the API operates in the background, allowing third-party developers to write programs that send data requests to OpenAI’s models and receive responses that those third-party programs can use.⁸

⁶ OpenAI, *ChatGPT*, <https://openai.com/chatgpt/overview/>.

⁷ “API” refers to an application programming interface. *See generally* OpenAI, *API Platform*, <https://openai.com/api/>.

⁸ OpenAI, *API Reference*, <https://platform.openai.com/docs/api-reference/introduction>.

Businesses typically use the API to power other services and applications, including in healthcare,⁹ finance,¹⁰ and translation¹¹ applications—among many others. For example, legal research services use OpenAI’s API to process legal research queries from their customers using OpenAI’s large language models,¹² and these API customers rely on OpenAI’s data retention policies when interfacing with their own customers.

Because data privacy and security are paramount to businesses, OpenAI highlights the myriad enterprise-grade features for “[s]ecurity and data privacy” that protect businesses who use the API, including:

- Business Associate Agreements (BAA) for HIPAA compliance
- Industry-leading encryption
- Multi-factor user authentication
- Private connections to cloud services
- Zero data retention policy by request
- Strict security and data protection standards (i.e., SOC 2 Type 2 compliance)¹³

Here, too, OpenAI makes detailed commitments to API customers regarding their data because, as OpenAI tells its customers, “[y]our data is your data.”¹⁴ In general, API customers direct the retention of their own records, not OpenAI. API customers can even select the specific physical location of the infrastructure OpenAI uses to facilitate the API service—underscoring the

⁹ OpenAI, *Healthify*, <https://openai.com/index/healthify/>.

¹⁰ OpenAI, *Endex Builds the Future of Financial Analysis, Powered by OpenAI’s Reasoning Models*, <https://openai.com/index/endex/>.

¹¹ OpenAI, *Minnesota’s Enterprise Translation Office Uses ChatGPT to Bridge Language Gaps* (Sept. 26, 2024), <https://openai.com/index/state-of-minnesota/>.

¹² OpenAI, *Hebbia’s Deep Research Automates 90% of Finance and Legal Work, Powered by OpenAI*, <https://openai.com/index/hebbia/>; OpenAI, *Harvey*, <https://openai.com/index/harvey/>.

¹³ OpenAI, *API Platform*, <https://openai.com/api/>.

¹⁴ OpenAI, *Data Controls in the OpenAI Platform*, <https://platform.openai.com/docs/guides/your-data>.

customer’s control over their data. *Id.* These commitments are ironclad. OpenAI retains API conversation data it stores pursuant to its 30-day retention period.¹⁵

B. News Plaintiffs’ Suits

This dispute implicates three of the lawsuits in this multidistrict litigation brought by the New York Times, eight regional newspapers, and the Center for Investigative Reporting (“CIR”) (collectively, “News Plaintiffs”): *New York Times v. Microsoft*, No. 1:23-cv-11195; *Daily News v. Microsoft*, No. 1:24-cv-03285; and *The Center for Investigative Reporting, Inc. v. OpenAI*, No. 1:24-cv-04872. Like all the cases in this MDL, News Plaintiffs allege that OpenAI “used [their] copyrighted works . . . to train [] large language models (LLMs).” *MDL ECF 1 at 2* (Transfer Order). But News Plaintiffs go further, also alleging that OpenAI infringed their copyrights because OpenAI’s users have allegedly deployed ChatGPT to access copies of their articles without paying for them. *See, e.g., NYT ECF 170 ¶ 180*. News Plaintiffs have offered no support for that allegation, *see infra* 13–14, even though OpenAI offered them the ability to test their claim by analyzing a statistically significant sample of the tens of billions of ChatGPT conversations that OpenAI retains in the ordinary course of business. Although OpenAI has been urging News Plaintiffs since last year to propose an approach to sample that retained data, they did not do so until late last month—almost *six months after* OpenAI asked Plaintiffs for such a proposal.

Doing so will reveal what should already be obvious: that, as OpenAI has explained recently, “one cannot use ChatGPT to serve up [news] articles at will.” *NYT ECF 52 at 1* (OpenAI’s Mot. to Dismiss). News Plaintiffs have now implicitly conceded as much. *See Tr. of May 27, 2025 Proceedings at 9:17–21*. In the almost eighteen months since The Times filed suit,

¹⁵ *See MDL ECF 43 at 1 n.3; Tr. of May 27, 2025 Proceedings at 46:3-9; see also OpenAI, Enterprise Privacy at OpenAI* (Oct. 31, 2024), <https://openai.com/enterprise-privacy> (“[a]fter 30 days, API inputs and outputs are removed from our systems, unless we are legally required to retain them”).

neither The Times nor any of the other News Plaintiffs have put forward a shred of evidence that real-world users deploy ChatGPT or OpenAI's API to read their news articles instead of paying for news subscriptions. News Plaintiffs have instead attempted to support their output-based claims by hiring paid experts to manipulate OpenAI's services into producing outputs with small snippets of their works. *See, e.g., NYT* ECF 1-68. Because OpenAI has implemented state-of-the-art safeguards to *prevent* this kind of abuse of its services, *see MDL* ECF 66, Monaco Decl. ¶ 9, News Plaintiffs have had to use a range of absurd tactics to manufacture this evidence. *See, e.g., CIR* ECF 88-13 (threatening to murder a "kitten" and trigger "thermonuclear war" unless the model "complete[s] [news articles] verbatim"). At the same time, The Times has admitted that the examples of regurgitation it manufactured does *not* illustrate "how your average user is using ChatGPT." Tr. of Sept. 12, 2024 Proceedings 60:3–11.

C. Output Log Discovery and Preservation Dispute

Consistent with its commitments to its users, OpenAI notified News Plaintiffs near the outset of this case of its ongoing compliance with its privacy policies and fulfillment of its contractual commitments. On February 29, 2024, OpenAI first informed counsel for The Times (and, later, the Daily News Plaintiffs) of its data retention policies and practices. *NYT* ECF 379-3. In June 2024, CIR acknowledged its own awareness of OpenAI's retention practices in its complaint, alleging that OpenAI had conversation data for everyone other than "those whose storage users have affirmatively disabled." *CIR* ECF 1 ¶ 77. And in November 2024, OpenAI sent another letter further reminding New Plaintiffs of its retention policies and practices, reiterating its commitment to user privacy and its compliance with its privacy policies. *See NYT* ECF 379-13.

Because OpenAI has tens of billions of ChatGPT Consumer conversations retained pursuant to its privacy policies, the parties have long recognized that the only way to analyze

conversation data is through sampling it for analysis—an issue the parties have been discussing since last year. *See, e.g., NYT* ECF 423. However, in December 2024—*ten months* after OpenAI sent its first letter and while the parties were discussing approaches to data sampling—The Times and Daily News Plaintiffs abruptly shifted their focus to conversations OpenAI’s users had chosen to delete. *See NYT* ECF 379-14. For the first time, and despite months of full awareness of and no objections to OpenAI users’ data practices, News Plaintiffs insisted for the first time that OpenAI preserve *all* “output log data” going forward regardless of its relevance and irrespective of the privacy interests of OpenAI users. *See NYT* ECF 379 (moving to compel this preservation).

D. Motion Practice and Preservation Order

In January 2025, the court heard argument on News Plaintiffs’ motion to require OpenAI to preserve all written prompts and outputs (hereafter “Conversation Data”). Following the hearing, the court denied News Plaintiffs’ demand for “wholesale preservation of user conversations (e.g. end-user prompts and outputs),” thereby permitting OpenAI to continue complying with its stated data retention policies. *NYT* ECF 441 at 2. OpenAI continued attempting to work with Plaintiffs on moving forward with sampling the billions of available conversations. *See, e.g., NYT* ECF 379-3.

Five months later, on May 12, 2025, News Plaintiffs re-raised this issue in a “supplemental letter,” in which they falsely accused OpenAI of deleting large batches of user data in response to litigation events. *MDL* ECF 27. Before OpenAI had an opportunity to respond to those unfounded accusations, the court ordered OpenAI to “preserve and segregate all output log data that would otherwise be deleted on a going forward basis until further order of the Court (in essence, the output log data that OpenAI has been destroying).” *MDL* ECF 33 at 2.

OpenAI quickly moved for reconsideration, detailing the significant burden and threat to user privacy posed by wholesale preservation and debunking News Plaintiffs’ baseless accusation

that OpenAI had intentionally deleted data in response to litigation events. *MDL ECF 40* at 3–4.¹⁶ (Tellingly, since then, News Plaintiffs have fallen silent on the matter, underscoring that these accusations were meritless all along.) OpenAI also pointed out (as it had before) that preservation of additional data was unnecessary because the data OpenAI had already retained—*i.e.*, tens of billions of user conversations—was more than sufficient to supply a statistically representative sample for analysis. *Id.* at 2.

On May 16, 2025, the court denied OpenAI’s motion for reconsideration without prejudice to renewal. *MDL ECF 42*. The court expressed concern that if people use ChatGPT for infringing purposes—*e.g.*, to “get around the [Times’s] pay wall”—they *might* be more likely to “delete all [their] searches” to cover their tracks. *Id.* at 4. At a subsequent hearing on May 27, 2025, the court explained that a chief concern was that, if users did so, honoring their requests to process those deletions might cause “[REDACTED].” Tr. of May 27, 2025 Proceedings at 4:13–20. Accordingly, in its May 16 order, the court requested “proposals” for “a means to determine whether the retained output logs marked for deletion differ in any material way from the retained logs.” *MDL ECF 42* at 5.

Shortly thereafter, OpenAI proposed exactly such a mechanism, which the News Plaintiffs refused to accept. *MDL ECF 68*, Nightingale Dawson Decl. ¶ 5. At the May 27, 2025 hearing, the court endorsed OpenAI’s proposal and instructed the parties to pursue it. *See* Tr. of May 27, 2025 Proceedings at 43:11–18. The court then issued an order directing OpenAI to “submit a proposed order regarding the specific sampling process, consistent with the discussion on the record, by June 6, 2025” and to “set out timelines for completion.” *MDL ECF 79*. Although the

¹⁶ Notwithstanding the burden on OpenAI and its users, OpenAI also took the steps that were immediately available to comply with the Preservation Order, resulting in the preservation of text prompts and outputs for ChatGPT Consumer and the API on a going-forward basis. *See MDL ECF 40* at 2; Tr. of May 27, 2025 Proceedings at 29:23–25.

court stated on the record that the Preservation Order will remain in place “[REDACTED]” and expressed an intent “[REDACTED],” Tr. of May 27, 2025 Proceedings at 51:14–16, the court declined to vacate the Preservation Order pending execution of the sampling process, *MDL ECF 79*.¹⁷

The court also declined to lift the Preservation Order as to API Conversation Data even though the stated bases for the Preservation Order—that users engaged in certain types of activity might be more likely to “delete all [their] searches”—do not apply to the API. *See infra* 14. Since API Conversation Data is subject to standard retention policies, API customers cannot “delete all [their] searches” based on the activity the customers (or their end-users) engaged in. The court nevertheless required OpenAI to continue preserving API Conversation Data as well. *See* Tr. of May 27, 2025 Proceedings at 51:21–24.

III. STANDARD OF REVIEW

Under Federal Rule of Civil Procedure 72(a), a district court must “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). A magistrate judge’s order is “contrary to law” if it “fails to apply or misapplies relevant statutes, case law, or rules of procedure.” *Coventry Capital U.S. LLC v. EEA Life Settlements Inc.*, 333 F.R.D. 60, 64 (S.D.N.Y. 2019) (citation omitted).

IV. ARGUMENT

The Federal Rules contemplate that a party take “reasonable steps” to preserve data that are proportional to the needs of the case. Fed. R. Civ. P. 37(e); *id.* advisory committee’s notes to 2015 amendment. Preservation is thus “evaluated on a standard of reasonableness, which is in

¹⁷ OpenAI is preparing its submission to the court regarding its proposed sampling process, as the court contemplated in its order at *MDL ECF 79*. OpenAI intends to work with News Plaintiffs and the court to ensure that the process moves forward expeditiously. Since the Preservation Order remains in place at present without a specific end date, OpenAI is submitting this Objection in parallel to ensure that both its rights and the rights of its users are protected.

turn informed by proportionality.” *Al Otro Lado, Inc. v. Nielsen*, 328 F.R.D. 408, 416 (S.D. Cal. 2018). Given that framework, courts “should not” compel parties to “preserve potentially relevant evidence . . . in the absence of some substantial basis for believing that it would serve a useful purpose.” *In re Application of Chevron Corp.*, 736 F. Supp. 2d 773, 787 (S.D.N.Y. 2010).

The Preservation Order is contrary to law for three reasons. First, notwithstanding the fact that OpenAI already preserves tens of billions of ChatGPT Consumer conversations and has offered to make that data available for sampling, the Order requires OpenAI to preserve even more data without any “substantial basis for believing that [doing so] would serve a useful purpose.” *Chevron*, 736 F. Supp. 2d at 787. Second, the Order is not proportional because any justification for ongoing preservation is far outweighed by the massive burden the Order imposes on both OpenAI and its users. Third, the Order was premised on a letter from News Plaintiffs that contained now-debunked accusations and material omissions.

OpenAI appreciates that the court has clarified both that (a) the Preservation Order is limited to what OpenAI has been able to address to date—preservation of ChatGPT Consumer and API Conversation Data on a going-forward basis—and (b) the Preservation Order will only remain in place for a limited (but still indefinite) period of time going forward. But because the Order remains in place (with no end date) and continues to prevent OpenAI from respecting its users’ privacy decisions without any just cause, OpenAI objects to the Order and hereby requests its vacatur.

A. Ongoing Preservation Will Not Serve A Useful Purpose

Preservation should not be ordered until, at a minimum, the party seeking such an order has established a need for the documents or data in question. *See Al Otro Lado*, 328 F.R.D. at 416; *see also Chevron*, 736 F. Supp. 2d at 787. That did not happen here.

To be clear, News Plaintiffs have already conceded that the data OpenAI retains in the ordinary course—*i.e.*, tens of billions of consumer ChatGPT conversations—is far too voluminous to analyze directly, and that the only way to proceed is by analyzing a sample of that data. *See, e.g., NYT ECF 423.* Instead of focusing on that exercise, they now demand that OpenAI suspend its longstanding data retention policies and ignore its users’ privacy decisions in order to preserve *even more* data. Yet they implicitly concede that at least 99.994% of the data they demand OpenAI preserve will not “implicate [their] claims” at all. *MDL ECF 68, Nightingale Dawson Decl. Ex. A.*

In fact, the *only* basis News Plaintiffs have proffered to justify their demand is their supposition that users who somehow allegedly derive ChatGPT outputs containing copies of News Plaintiffs’ articles will “cover their tracks” by deleting those outputs—which, in turn, will exclude these supposedly infringing outputs from the universe of data they plan to sample. As a threshold matter: the notion that OpenAI’s users have derived ChatGPT outputs containing copies of News Plaintiffs’ articles is completely unsupported by the pleadings and the available evidence. News Plaintiffs have been litigating this case for well over a year, and have had full access to ChatGPT throughout that period—yet at no point during the briefing of this issue have News Plaintiffs put forward any evidence that an ordinary user could use ChatGPT to obtain copies of their articles. Indeed, they have now effectively admitted that it is “difficult” if not impossible to do so. *See MDL ECF 43 at 9.* In other words, the factual premise of News Plaintiffs’ argument—*i.e.*, that ChatGPT allows users to “evade publishers’ paywalls,” *MDL ECF 43 at 2*—is simply incorrect and, more to the point, entirely unsupported.

Moreover, the notion that ChatGPT users will selectively delete output logs to “cover their tracks,” as News Plaintiffs theorize, finds no support in the record. *MDL ECF 43 at 2* (speculating

that “users who were using ChatGPT to evade publishers’ paywalls were also more likely to ask OpenAI to delete their output data, and [that] the prevalence of harmful outputs is [therefore] higher in the destroyed data”). At no point during the five months of litigation of this issue have News Plaintiffs proffered any evidence to support this theory, notwithstanding OpenAI’s production of over 110,000 documents in this case. Speculation alone is obviously insufficient to provide the requisite “substantial basis for believing” that preservation “would serve a useful purpose.” *Chevron*, 736 F. Supp. 2d at 787; *see also Griffith v. TikTok, Inc.*, No. 5:23-cv-00964, 2024 WL 5279224, at *8 (C.D. Cal. Dec. 24, 2024) (“Plaintiffs have not shown any basis for the Court to infer that the deleted data contained evidence necessary to support their claims that is materially different from the data in the sample.”).

News Plaintiffs’ justification makes even less sense as applied to data from OpenAI’s API service. News Plaintiffs’ *only* basis for ongoing preservation—that users delete specific ChatGPT conversations to “cover their tracks”—is inapplicable to API users as they cannot request that OpenAI delete specific API logs subject to OpenAI’s standard 30-day retention policy. Further, the API is not a consumer-facing tool: it is a mechanism for third-party software developers to write computer code that incorporates OpenAI’s models into other applications, like legal research tools, translation apps, and the like. *See supra* 5–6. News Plaintiffs have never explained why they think any such developer would utilize the API to read their news articles. Because News Plaintiffs have proffered no other justification for demanding preservation of API logs beyond the 30-day retention of all stored API logs pursuant to OpenAI’s existing policies, and because those logs have no conceivable relevance to their claims, the Order is wholly unjustified as applied to the API as well.

B. The Order Is Not Proportional to the Needs of the Case

A key “factor in evaluating the reasonableness of preservation efforts is proportionality.” Fed. R. Civ. P. 37 advisory committee’s notes to 2015 amendment; *see also Al Otro Lado*, 328 F.R.D. at 416 (“Preservation is evaluated on a standard of reasonableness, which is in turn informed by proportionality.”). An evaluation of proportionality in this context necessarily requires balancing the burden of preservation against the contents of the documents sought to be preserved. *See Boudreau v. Petit*, No. 17-cv-301, 2025 WL 553975, at *2 (D.R.I. Feb. 19, 2025).

The Preservation Order here does not reflect any such balancing and is wholly disproportionate to the needs of the case. The Order imposes a substantial burden on OpenAI and its users. On its face, the Order would require OpenAI to make significant changes to its data infrastructure, requiring up to [REDACTED] of dedicated engineering time, along with additional costs [REDACTED] of dollars. *MDL ECF 66*, Monaco Decl. ¶ 8. Taking those steps would also damage OpenAI’s relationships with its users, who rely on OpenAI’s commitment to data privacy when using OpenAI’s services. Like any cache of private information in the digital age, ChatGPT conversations range “from the mundane to the intimate.” *Riley v. California*, 573 U.S. 373, 395 (2014). Users’ ChatGPT conversations include personal information about relationships with loved ones, confidential business plans, proprietary business information and trade secrets—and many other kinds of information that users have strong personal or professional reasons for safeguarding. The assurance that OpenAI will process deletions of this information if users choose to delete it is absolutely essential to users’ ability to engage openly with OpenAI’s services. *MDL ECF 67*, Kechida Decl. ¶ 3. This is precisely why members of the public immediately voiced concern over the Preservation Order. *See MDL ECF 68*, Nightingale Dawson Decl. Exs. C–L; *see*

also MDL ECF 49, Mot. to Intervene (objecting to the News Plaintiffs’ “overreaching into the private affairs of third parties with no stake in this litigation”).¹⁸

An order requiring retention of this information—regardless of users’ decisions and without substantial basis—not only imposes a technical burden, but also undermines the ongoing trust that users place in OpenAI and raises “serious privacy concerns.” *Gonzales v. Google, Inc.*, 234 F.R.D. 674, 687 (N.D. Cal. 2006). A court should weigh those concerns *before* issuing a preservation order. *See Gill v. Gulfstream Park Racing Ass’n., Inc.*, 399 F.3d 391, 402 (1st Cir. 2005) (courts must “balance[]” “[c]onsiderations of the public interest, the need for confidentiality, and privacy interests” when determining discoverability); *Al Otro Lado*, 328 F.R.D. at 416 (“[T]he scope of a party’s duty to preserve is the same as the scope of discovery articulated in Rule 26(b)(1).”). That did not happen here. And once those concerns are properly accounted for, it is clear that the Preservation Order is *not* proportionate to the needs of the case.

C. The Order Rests on False and Disproven Factual Assumptions

Finally, the Order was issued in direct response to—and was seemingly premised on—a letter from the News Plaintiffs that both misrepresented and omitted critical, relevant facts. First, the Order was based on News Plaintiffs’ blatant misrepresentation of OpenAI’s data retention practices. Notably, this was not the first time News Plaintiffs have attempted to force OpenAI to preserve “all of its output log data” in January of this year, arguing that this was necessary to preserve relevant evidence. *NYT* ECF 379 at 3. The court below rejected that request, including by recognizing the substantial “privacy and technological considerations implicated in Plaintiffs’ request.” *NYT* ECF 441 at 2. Rather than accept that result, News Plaintiffs tried a different tack

¹⁸ The court denied the motion to intervene on procedural grounds. *MDL* ECF 76.

in May 2025: filing a letter falsely accusing OpenAI of “destroying” Conversation Data in response to litigation events.

The Preservation Order—which directed OpenAI to “preserve and segregate” the “output log data that OpenAI has been destroying,” *MDL ECF 33* at 2—was issued the day after Plaintiffs filed that letter on the docket. As a result, OpenAI never had the opportunity to explain, prior to the Order’s issuance, that the accusations on which the Order was based were entirely fabricated. To be clear—and as OpenAI has now explained in a sworn declaration, *MDL ECF 66*, Monaco Decl. ¶¶ 10–14—OpenAI did not “destroy” any data, and certainly did not delete any data in response to litigation events. The Order appears to have incorrectly assumed the contrary, *see MDL ECF 33* at 2 (assuming that “OpenAI has been destroying” data), and thus rests on a faulty and now-debunked premise.

Second, News Plaintiffs’ May letter, and thus the Preservation Order, also did not address the fact that News Plaintiffs have long known about OpenAI’s privacy policies and compliance therewith, and did not object for almost a year (or more). While News Plaintiffs have repeatedly mischaracterized the record on this point, *see, e.g. NYT ECF 379; NYT ECF 539-1; NYT ECF 560*, it is beyond dispute that OpenAI informed The New York Times’s counsel (also counsel for the Daily News plaintiffs) of its data practices in *February 2024*. *NYT ECF 379-3* (letter explaining data practices for “ChatGPT” and “API” and explaining that “it would be inappropriate, overly burdensome, not proportional, and not feasible to place all ChatGPT inputs and output on indefinite retention”). The New York Times did not respond and did not object to OpenAI’s continued privacy policy compliance until *ten months later*—and only after OpenAI sent yet another similar letter in *November 2024*. *NYT ECF 379-13*. And CIR *never* raised an issue regarding OpenAI’s privacy policies and retention practices until last month, despite knowing of

them since at least June 2024. *See supra* 8. News Plaintiffs’ seeming deliberate omission of this critical history, which itself amounts to an implicit acknowledgement by them that wholesale preservation is “inappropriate,” *NYT* ECF 379-3, likewise resulted in an Order that rests on erroneous premise, rendering it without “substantial basis” and thus contrary to law.

V. CONCLUSION

OpenAI respectfully requests that the Court grant OpenAI’s objection and vacate the Preservation Order at *MDL* ECF 33.

Dated: June 3, 2025

/s/ Elana Nightingale Dawson
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

Sarang V. Damle
sy.damle@lw.com
Elana Nightingale Dawson (*pro hac vice*)
elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Rachel R. Blitzler
rachel.blitzler@lw.com
Herman H. Yue
herman.yue@lw.com

Luke A. Budiardjo
luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

Allison S. Blanco (*pro hac vice*)
allison.blanco@lw.com
650 Town Center Drive 20th Floor
Costa Mesa, CA 92626
Telephone: 714.540.1235

Attorneys for OpenAI Defendants

/s/ Joseph C. Gratz
MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)*
jgratz@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

Carolyn M. Homer (*pro hac vice*)
cmhomer@mofo.com
2100 L Street, NW, Suite 900
Washington, DC 20037
Telephone: 202.650.4597

Eric K. Nikolaidis
enikolaides@mofo.com
Emily C. Wood
ewood@mofo.com
Jocelyn E. Greer
jgreer@mofo.com
250 West 55th Street
New York, NY 10019
Telephone: 212.468.8000

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.

/s/ Edward A. Bayley

KEKER, VAN NEST & PETERS LLP

Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com

Paven Malhotra
pmalhotra@keker.com

Michelle S. Ybarra (*pro hac vice*)
mybarra@keker.com

Nicholas S. Goldberg (*pro hac vice*)
ngoldberg@keker.com

Thomas E. Gorman (*pro hac vice*)
tgorman@keker.com

Katie Lynn Joyce (*pro hac vice*)
kjoyce@keker.com

Sarah Salomon (*pro hac vice*)
ssalomon@keker.com

R. James Slaughter (*pro hac vice*)
rslaughter@keker.com

Christopher S. Sun (*pro hac vice*)
csun@keker.com

Andrew F. Dawson (*pro hac vice*)
adawson@keker.com

Andrew S. Bruns (*pro hac vice*)
abrun@keker.com

Edward A. Bayley (*pro hac vice*)*
ebayley@keker.com

Ryan K. Wong (*pro hac vice*)
rwong@keker.com

Spencer McManus (*pro hac vice* pending)
smcmanus@keker.com

Olivia C. Malone (*pro hac vice* pending)
omalone@keker.com

633 Battery St.
San Francisco, CA 94111
Telephone: 415.391.5400

Attorneys for OpenAI Defendants