

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

IN RE:

OpenAI, INC.,

25-md-3143 (SHS) (OTW)

COPYRIGHT INFRINGEMENT LITIGATION

This Document Relates To:

News Cases

1:23-cv-11195-SHS-OTW

1:24-cv-01515-SHS-OTW

1:24-cv-03285-SHS-OTW

1:24-cv-04872-SHS-OTW

1:25-cv-04315-SHS-OTW

**MEMORANDUM OF LAW IN SUPPORT OF NEWS PLAINTIFFS' MOTION FOR
SANCTIONS AGAINST OPENAI**

Table of Contents

I.	FACTUAL BACKGROUND	5
A.	OpenAI Misrepresents Its Ability to Search Its Training Datasets.....	6
B.	OpenAI Mischaracterizes its Reservoirs of Already De-Identified and Searchable Output Logs, Forcing the Court and News Plaintiffs to Accept the 20 Million Sample	9
C.	OpenAI Obstructs the ChatGPT Output Log Sampling Process.	11
1.	OpenAI Deleted or Compressed Billions of ChatGPT Conversations, Thereby Rendering Them Unsearchable.....	12
2.	OpenAI’s Substitution of Millions of Plaintiffs’ Requested Logs Further Skewed the 20 Million Sample	17
3.	OpenAI Reneged On Its Agreement to Produce the 20 Million Logs, Then, When Ordered to Do So, Over-Redacted Them So as to Render Them “Unusable.”.....	18
D.	OpenAI Obstructs Mr. Monaco’s First Deposition.....	20
E.	Mr. Monaco’s Second Deposition Reveals OpenAI’s Search Capabilities and Already-Conducted Searches for News Plaintiffs’ Content.....	21
F.	OpenAI Claims that it Already “Produced” the ChatGPT Dataset it Had Been Refusing to Produce for Months.....	24
II.	LEGAL FRAMEWORK	24
III.	ARGUMENT	26
A.	OpenAI Should be Precluded From Relying on the 20 Million Sample Pursuant to Rule 37(b) and (e)(1) and this Court’s Inherent Authority.....	27
1.	Preclusion Is an Appropriate Remedy Under Rule 37(b) Due to OpenAI’s Violation of This Court’s Orders.	27
2.	Preclusion Is an Appropriate Remedy Under Rule 37(e)(1) Due to OpenAI’s Failure To Preserve Output Logs.	31
B.	This Court Should Find That the ChatGPT Output Logs Include (Or Would Have Shown if Properly Produced in Discovery) Substantial and Systematic Grounding on and Regurgitation of News Plaintiffs’	

Copyrighted Material Into User Outputs, and OpenAI Should Be
Prohibited From Arguing to the Contrary..... 33

C. This Court May Also Impose Each of the Requested Remedies
Pursuant to Rule 37(e)(2) and Should Instruct the Jury That OpenAI
Deleted Billions of Output Logs. 38

D. This Court Should Award News Plaintiffs Attorneys’ Fees in
Connection with Their Motions and Meet and Confers Regarding
Project Giraffe, Depositions of Mr. Monaco, and Time Their Lawyers
and Experts Spent Reviewing Training and Output Material in the
“Sandbox.” 39

E. The Court May Also Impose Each of the Requested Remedies Pursuant
to its Inherent Authority..... 42

IV. CONCLUSION..... 43

TABLE OF AUTHORITIES

Cases

<i>Agiwal v. Mid Island Mortg. Corp.</i> , 555 F.3d 298 (2d Cir. 2009)	26, 27
<i>Arista Records, Inc. v. Mp3Board, Inc.</i> , No. 00-cv-4660, 2002 WL 1997918 (S.D.N.Y. Aug. 29, 2002)	32
<i>Brown v. City of New York</i> , No. 15-cv-4488, 2018 WL 3193208 (E.D.N.Y. June 28, 2018)	29
<i>CBF Industria de Gusa S/A v. AMCI Holdings, Inc.</i> , No. 13-cv-2581, 2021 WL 4190628 (S.D.N.Y. Aug. 18, 2021)	38
<i>Chambers v. NASCO, Inc.</i> , 501 U.S. 32 (1991)	26, 39, 42
<i>Charlestown Cap. Advisors, LLC v. Acero Junction, Inc.</i> , 337 F.R.D. 47 (S.D.N.Y. 2020)	32
<i>Cine Forty-Second St. Theatre Corp. v. Allied Artists Pictures Corp.</i> , 602 F.2d 1062 (2d Cir. 1979)	37
<i>Doubleline Capital LP v. Odebrecht Fin., Ltd.</i> , No. 17-cv-4576, 2021 WL 1191527 (S.D.N.Y. Mar. 30, 2021)	33
<i>Eid v. Koninklijke Luchtvaart Maatschappij N.V.</i> , 310 F.R.D. 226 (S.D.N.Y. 2015)	41
<i>Fashion Exch. LLC v. Hybrid Promotions, LLC</i> , No. 14-cv-1254, 2021 WL 1172265 (S.D.N.Y. Mar. 29, 2021)	38
<i>Fendi Adele v. Filene's Basement, Inc.</i> , No. 06-cv-244 (RMB), 2009 WL 855955 (S.D.N.Y. Mar. 24, 2009)	37
<i>Fujitsu Ltd. v. Fed. Express Corp.</i> , 247 F.3d 423 (2d Cir. 2001)	31
<i>Hinds v. Cnty. of Westchester</i> , No. 11-cv-7265, 2020 WL 7046843 (S.D.N.Y. Dec. 1, 2020)	26
<i>Hoffer v. Tellone</i> , 128 F.4th 433 (2d Cir. 2025)	38
<i>In re Keurig Green Mountain Single-Serve Coffee Antitrust Litig.</i> , 341 F.R.D. 474 (S.D.N.Y. 2022)	36

<i>J.C. v. Zimmerman</i> , 150 F.4th 136 (2d Cir. 2025)	25
<i>Joint Stock Co. Channel One Russia Worldwide v. Infomir LLC</i> , No. 16-cv-1318 (GBD), 2019 WL 4727537 (S.D.N.Y. Sept. 26, 2019), <i>aff'd sub nom. Joint Stock Co. "Channel One Russia Worldwide" v. Infomir LLC</i> , No. 16-cv-1318 (GBD), 2020 WL 1479018 (S.D.N.Y. Mar. 26, 2020).....	30, 31
<i>Karsch v. Blink Health Ltd.</i> , 17-cv-3880 (BCM), 2019 WL 2708125 (S.D.N.Y. June 20, 2019)	32
<i>Kassenoff v. Kassenoff</i> , No. 22-cv-2162, 2024 WL 291225 (S.D.N.Y. Jan. 25, 2024)	26
<i>Klein v. Torrey Point Grp., LLC</i> , 979 F. Supp. 2d 417 (S.D.N.Y. 2013)	42
<i>Medcenter Holdings Inc. v. Web MD Health Corp.</i> , 734 F. Supp. 3d 303 (S.D.N.Y. 2024)	31
<i>Nat'l Hockey League v. Metro. Hockey Club, Inc.</i> , 427 U.S. 639 (1976).....	42
<i>New York State Nat. Org. for Women v. Cuomo</i> , 1997 WL 671610 (S.D.N.Y. Oct. 28, 1997).....	29
<i>Novak v. Wolpoff & Abramson LLP</i> , 536 F.3d 175 (2d Cir. 2008)	39, 42
<i>Oz v. Lorowitz</i> , No. 09-cv-5532, 2011 WL 803077 (S.D.N.Y. Mar. 7, 2011).....	36
<i>Patterson v. Balsamico</i> , 440 F.3d 104 (2d Cir. 2006)	25, 28
<i>Polaroid Corp. v. Casselman</i> , 213 F. Supp. 379 (S.D.N.Y. 1962)	34
<i>Reilly v. Natwest Mkts. Grp. Inc.</i> , 181 F.3d 253 (2d Cir. 1999)	24
<i>Residential Funding Corp. v. DeGeorge Fin. Corp.</i> , 306 F.3d 99 (2d Cir. 2002)	40
<i>Rosbach v. Montefiore Med. Ctr.</i> , 81 F.4th 124 (2d Cir. 2023)	26

<i>S.E.C. v. Rajaratnam</i> , 622 F.3d 159 (2d Cir. 2010)	29
<i>Shcherbakovskiy v. Da Capo Al Fine, Ltd.</i> , 490 F.3d 130 (2d Cir. 2007)	26
<i>Skyline Steel, LLC v. PilePro, LLC</i> , 101 F. Supp. 3d 394 (S.D.N.Y. 2015)	32
<i>Softel, Inc. v. Dragon Medic. & Sci. Commc'ns, Inc.</i> , 118 F.3d 955 (2d Cir. 1997)	25
<i>Zubulake v. UBS Warburg LLC</i> , 220 F.R.D. 212 (S.D.N.Y. 2003)	12

Rules

Fed. R. Civ. P. 26.....	25, 35
Fed. R. Civ. P. 33.....	25
Fed. R. Civ. P. 34.....	25
Fed. R. Civ. P. 37.....	passim
Fed. R. Civ. P.30.....	28

This is a case about copying. There is no question that it happened. Nor should there be one about what was copied, how often, or to what end. The evidence is in OpenAI’s training datasets and ChatGPT output logs. But instead of just producing that evidence at the start of the case and focusing on the merits of its fair use defense, OpenAI chose obstruction. It told News Plaintiffs,¹ and the Court, that it could not search for evidence that it copied their works to train its models, or again to ground their outputs, or yet again in the outputs themselves. None of this was true. OpenAI had done all these things—starting even before the first News Plaintiff filed suit. Only a compelled re-deposition of an ill-prepared witness exposed the facts:

OpenAI’s Misrepresentation to the Court	OpenAI’s (Second) R.30(b)(6) Deposition
September 2024 OpenAI does “not currently have tools to search [the training datasets] efficiently”; “it is not a situation where we could do it very quickly and it would take them a very long time, <i>it would also take us a very long time.</i>”²	April 2026 
May 2025 “OpenAI’s current infrastructure also <i>does not include any mechanism that could search a large dataset of output logs</i> like decompressed Conversation Data for specific content. OpenAI would need to design a custom system...Designing a proof-of-concept for such a system would take at least	April 2026 

¹ News Plaintiffs include The New York Times (“The Times”), Daily News Plaintiffs, the Center for Investigative Reporting (“CIR”), the Intercept, and Ziff Davis Plaintiffs. Class Plaintiffs intend to file a separate spoliation motion prior to the filing of summary judgment motions related to OpenAI’s destruction of books-related data.

² See Declaration of Jennifer Maisel in Support of News Plaintiffs’ Motion for Sanctions (“Maisel Decl.”), Ex. 1 (Sept. 12, 2024, Hearing Tr.) at 7.

³ Maisel Decl., Ex. 2 (Apr. 8, 2026, V. Monaco 30(b)(6) Dep. Tr.) at 220:5-10 (emphasis added).

videoconference, none of which would have been required had OpenAI been forthcoming from the beginning.⁸ Instead, its two-year effort to hide its ability to search its training data and output logs for News Plaintiffs' copyrighted content surfaced only during the second deposition of its five-time declarant and Rule 30(b)(6) witness Vincent Monaco, whom the Court ordered re-deposed for lack of preparation and OpenAI's obstruction of his first deposition. Dkt. [1492](#). OpenAI had repeatedly claimed that it could not conduct such searches to respond to News Plaintiffs' discovery requests, but Monaco admitted [REDACTED] [REDACTED] OpenAI's concealment of this fact withheld highly relevant evidence, prolonged discovery, inflated expenses, and burdened the Court.

And not inadvertently: its corporate representative and at least one of its outside counsel knew. After submitting multiple inaccurate declarations to this Court, further discovery from OpenAI's own 30(b)(6) witness finally revealed [REDACTED]

[REDACTED]
[REDACTED]⁹

OpenAI's sanctionable acts do not end with its misrepresentations. Rather, for over a year after these lawsuits were filed, OpenAI failed to preserve materials related to ChatGPT's output of News Plaintiffs' copyrighted content, including materials News Plaintiffs sought in their earliest

⁸News Plaintiffs do not suggest that OpenAI's attorneys were all aware that their client had already [REDACTED] or that every inaccurate statement they made to News Plaintiffs or this Court was intentionally misleading. But as explained herein, discovery has shown [REDACTED]. See *infra* at Section I(E).

⁹ As Mr. Monaco eventually testified, OpenAI [REDACTED] [REDACTED] "Ex. 2 at 253:23-254:5. He also testified that [REDACTED] *Id.* at 224:22-225:18; 225:22-226:5; 227:[3-10](#).

discovery requests. Indeed, OpenAI’s deletions continued even after this Court’s Order directing OpenAI to preserve and segregate these very materials. Dkt. [33](#).

OpenAI’s misconduct did not just cause delays and costs, although both were substantial. OpenAI’s actions impeded News Plaintiffs’ ability to ascertain the extent to which OpenAI scrapes, trains and grounds on, and outputs News Plaintiffs’ content—which further obstructed News Plaintiffs from developing these themes fully in discovery. OpenAI cannot have failed to understand and intend that its misconduct should produce a record that distorts the evidence in its favor and prejudices News Plaintiffs.

News Plaintiffs accordingly ask this Court to impose sanctions on OpenAI pursuant to Fed. R. Civ. P. 37(b), (c), (d), and (e), and this Court’s inherent authority. The requested sanctions focus on the issues to which the discovery was directed—i.e., OpenAI’s acquisition and use of News Plaintiffs’ copyrighted works and ChatGPT’s regurgitation of those materials in response to user prompts. Accordingly, News Plaintiffs seek the following sanctions:

1. OpenAI should be prohibited from relying on the 20 million “sample” ChatGPT output logs for any purpose.
2. This Court should find that ChatGPT’s output logs include (or would have shown if properly produced in discovery) substantial and systematic grounding on and regurgitation of News Plaintiffs’ copyrighted material in outputs.
3. OpenAI should be prohibited from arguing that ChatGPT’s output logs do not include (or would not have shown if properly produced in discovery) substantial and systematic grounding on and regurgitation of News Plaintiffs’ copyrighted material into user outputs.
4. This Court should instruct the jury on these findings and their binding effect.

5. The Court should award News Plaintiffs their attorneys' fees, expert fees, and costs associated with their efforts to secure the improperly withheld discovery, including through on-site inspections and analyses.

I. FACTUAL BACKGROUND

From the outset of this case, two key issues have been: (1) how much of News Plaintiffs' copyrighted content did OpenAI take and use to train its LLMs, and (2) to what extent do the LLMs "regurgitate" that content in response to users' prompts. OpenAI should have been gathering that information for production from the very beginning of these actions, and certainly in response to News Plaintiffs' February 2024 initial discovery requests focused on these issues. *See* Maisel Decl., Ex. 4 (Times Feb. 23, 2024 First Set of Requests for Production to OpenAI) at 6 (RFP No 1 seeking "[d]ocuments concerning the use of Times Content for training or fine tuning Your Text Generation AI Models"); as 7 (RFP No. 8 seeking "[d]ocuments sufficient to show to what extent and in what form Times Content resides in, is stored in, is copied by, or otherwise exists within the Training Datasets and/or the Text Generation AI Models"); and at 8 (RFP No. 14 seeking "complete listing of every data sample in the Training Datasets"). We now know, because OpenAI's designated corporate representative Mr. Monaco has testified to it under oath, that OpenAI possessed [REDACTED]. Specifically, [REDACTED]

[REDACTED].¹⁰ But rather than producing these analyses and the underlying data, OpenAI concealed their existence and told News Plaintiffs and the Court that it was incapable of searching and preserving its training data and

¹⁰ News Plaintiffs moved to compel OpenAI to produce or identify on a privilege log any such evaluations and related analyses that post-date the litigation, Dkt. [1520](#). The Court denied this request without prejudice, Dkt. [1551](#).

output logs in this way—all the while accusing News Plaintiffs of invading user privacy, assuring the public that any regurgitation of News Plaintiffs’ content in ChatGPT was “a rare bug,”¹¹ and continuing to delete output logs even after this Court’s Preservation Order.

This information finally came to light six weeks after fact discovery ended, when OpenAI’s 30(b)(6) witness on these topics (and five-time declarant before this Court), Mr. Monaco, was re-deposed following this Court’s Order finding that he had been ill-prepared for and wrongly precluded from answering relevant questions at his first deposition. Dkt. [1492](#) at 5, 8. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] Immediately after News Plaintiffs filed their lawsuits,

[REDACTED],

but—as discussed below—OpenAI intentionally hid its discovery capabilities from News Plaintiffs and the Court for two years. The Maisel Declaration attached hereto includes an appendix containing a chronology of the pertinent events.

A. OpenAI Misrepresents Its Ability to Search Its Training Datasets.

On December 27, 2023, The Times filed a detailed complaint alleging that Defendants infringed its copyrights by copying millions of human-authored Times works for commercial gain. On February 23, 2024, The Times served discovery requests directed to the improper acquisition and inclusion of News Plaintiffs’ copyrighted materials in OpenAI’s training data and ChatGPT outputs. *See* Ex. 4. For example, The Times asked OpenAI to produce “[d]ocuments

¹¹ Maisel Decl., Ex. 5 (Jan. 8, 2024 “OpenAI and journalism” blog post).

concerning the use of Times Content for training or fine tuning Your Text Generation AI Models.”¹² *Id.* at 6 (RFP No. 1). OpenAI responded that it would “make available for inspection, pursuant to an inspection protocol to be negotiated between the parties, the pretraining data for models used for ChatGPT that it locates after a reasonable search.” Maisel Decl., Ex. 6 (OpenAI’s Responses and Objections to The Times First Set of RFPs) at 8.

The parties spent the next five months engaged in numerous meet and confers and exhaustive email correspondence. On February 23, 2024, Class Plaintiffs moved to compel responses to Requests for Admission seeking “whether each of Plaintiffs’ works were included in OpenAI’s ‘training’ dataset and used to ‘train’ its GPT models.” *Authors Guild* Dkt. [78](#).¹³ At this Court’s September 12, 2024 hearing, OpenAI represented that it did “*not currently have tools to search [the training datasets] efficiently*” and “*it is not a situation where we could do it very quickly and it would take them a very long time, it would also take us a very long time.*” Ex. 1 at 7:17-18 (emphasis added). OpenAI proposed setting up a “sandbox”—what the Court described as “a very, very tightly controlled environment”—located in a single room in San Francisco (where neither of the law firms representing The Times and the Daily News Plaintiffs have offices), in which OpenAI would provide laptops for Plaintiffs to inspect the training datasets. *Id.* at 6:8-12. Taking OpenAI at its word that it had not itself searched for this information, and that doing so would be as cumbersome for OpenAI as for Plaintiffs, the Court ordered the creation of said “sandbox” and further conferrals on how to best equip Plaintiffs to run searches efficiently there. *Id.* at 6:15-21, 13:19-23 (Court questioning why it would be “more efficient, more practical, and

¹² The Daily News Plaintiffs served similar discovery on June 24, 2024. Maisel Decl., Ex 7.

¹³ News Plaintiffs also served requests for admission [REDACTED]

save resources” for plaintiffs to conduct searches themselves), *id.* 15:9-12; *see also Times Dkt. 304* at 4.¹⁴

OpenAI made operating in this restrictive environment even more difficult than News Plaintiffs anticipated, requiring dozens of emails and calls with OpenAI, to say nothing of numerous cross-country trips by both News Plaintiffs’ counsel and their experts and days spent locked in the sandbox. As Plaintiffs explained to the Court on November 2, 2024: “In the first week of the inspection alone, Plaintiffs experienced nearly a dozen disruptions to the inspection environment . . . OpenAI has refused to install some software packages provided in advance by News Plaintiffs . . . and shut down ongoing searches that the News Plaintiffs had initiated.” *See Times Dkt. 197* at 1-2; Maisel Decl., Ex. 53. In a subsequent update on November 20, Plaintiffs informed the Court that OpenAI engineers had erased “all of News Plaintiffs’ programs and search result data stored on one of the dedicated virtual machines,” and Plaintiffs reiterated that “OpenAI is in the best position to search its own datasets for the News Plaintiffs’ works using its own tools and equipment.” *Times Dkt. 328* at 2. News Plaintiffs were then forced to spend another three months negotiating searches for OpenAI to run. *Times Dkt. 462* at 2.

During this process, OpenAI withheld from the Court and News Plaintiffs that it had previously performed multiple searches for News Plaintiffs’ content. OpenAI first referenced Project Giraffe in a June 2024 attachment to an email, which included “Giraffe” as an ESI search term in a document listing term hits. Maisel Decl., Ex. 10 (June 3, 2024 OAI NYT Search Terms). On February 12, 2025, News Plaintiffs wrote to OpenAI identifying datasets in OpenAI’s Project Giraffe document and source code directory [REDACTED]

¹⁴ Citations to “*Times Dkt.*” refer to docket entries in *The New York Times*, Case No. 1:23-cv-11195.

[REDACTED] Maisel Decl., Ex. 11 at 2-3 (Feb. 12, 2025 Letter from Z. Savage). As Plaintiffs confirmed at Mr. Monaco’s second deposition, [REDACTED] [REDACTED] (Ex. 2 at 223:8-224:3), as well as two [REDACTED]” versions of those datasets where OpenAI [REDACTED] [REDACTED] (Ex. 2 at 216:17-217:21).

Following further meet and confers, in June 2025, OpenAI produced the [REDACTED]” datasets, but not the others. Of course, News Plaintiffs had no way to evaluate the datasets with confidence, or understand what they were used for, without a straight answer from OpenAI. *See* Maisel Decl., Ex. 12 (News Plaintiffs’ July 31, 2025 30(b)(6) Notice to OpenAI) at ¶¶ 71-76; *Times* Dkt. [560](#) at 10 (motion to compel seeking, *inter alia*, OpenAI 30(b)(6) witness in connection with output log data). But OpenAI refused to provide straight answers and did not produce a witness that was prepared and willing to testify about these datasets until Mr. Monaco’s Court-ordered second deposition in April 2026.

B. OpenAI Mischaracterizes its Reservoirs of Already De-Identified and Searchable Output Logs, Forcing the Court and News Plaintiffs to Accept the 20 Million Sample

In May and June 2024, News Plaintiffs served requests for production directed to the outputs of the LLMs (Maisel Decl. Exhs. 7, 13) as well as requests for inspection seeking “[q]uery, session, and chat logs related to The Times’s Content reflecting user sessions on [OpenAI’s] Generative AI Products and Services (‘Session Data’), including, for each session, user queries paired with responses to those queries,” Maisel Decl., Exhs. 14 at 2; 15 at 2. In response, OpenAI represented to News Plaintiffs, and then to this Court, that OpenAI could not readily search its ChatGPT output data, and the parties would therefore instead need to sample the data in the

sandbox. *See, e.g.*, Maisel Decl., Ex. 16 (June 26, 2025 Hr’g Tr.) at 284:8-15. (Counsel for OpenAI: “[W]e are working closely with Judge Wang to figure out how to sample those [chats]. Right? That has to be sampled. We can’t go through and – that’s not something that can be searched.”).

Not once did OpenAI disclose the existence [REDACTED]

[REDACTED] To the contrary, for more than two years, OpenAI repeatedly misled News Plaintiffs and this Court as to its discovery capabilities. *Times* Dkt. 462 (Feb. 13, 2025 Discovery Dispute Chart) at 4 (“OpenAI explained why, based on its investigation to date, the specific [Giraffe] tool Plaintiffs identified did not function in a way that aligned with the request Plaintiffs were making.”); Maisel Decl., Ex. 17 (Mar. 19, 2025 Letter from Counsel for OpenAI) (“OpenAI does not have an existing tool or mechanism that can search for ‘News Plaintiffs’ intellectual property.”).

On March 25, 2025, News Plaintiffs asked OpenAI to produce files corresponding to the [REDACTED] (Maisel Decl., Ex. 18 at 1-2) which, as Plaintiffs later learned at Mr. Monaco’s deposition, include an [REDACTED]

[REDACTED] Ex. 2 at 241:4-242:25. On June 4, 2025, Plaintiffs specifically asked OpenAI to produce a dataset called simply [REDACTED] which, again, Plaintiffs later learned included the 78 million de-identified ChatGPT conversations OpenAI [REDACTED]

[REDACTED] Maisel Decl., Ex. 19 (June 5, 2025 K. Logan Email to OpenAI).

When OpenAI refused to produce these datasets, on June 10, 2025, News Plaintiffs moved to compel. Dkt. [155](#). OpenAI answered that it “ha[d] not refused to produce anything,” and that News Plaintiffs had mischaracterized the datasets. Dkt. [185](#) at 1. The following June 23, 2025, email exchange is illustrative:

[News Plaintiffs]: “Giraffe [REDACTED] Dataset – OpenAI’s opposition to our motion represents that this dataset comprises [REDACTED] but does not further disclaim relevance based on *how* this dataset was used. Can you provide further information as to what this dataset was used for? Specifically, can you confirm whether this dataset was subsequently analyzed to identify instances of infringement? Given that this dataset is [REDACTED]. For example, looking at OPCO NEWS 0904663 we understand that OpenAI has a [REDACTED]. See also OPCO_NEWS_0943396.”

[OpenAI]: “*The basis for Plaintiffs motion to compel this dataset was that plaintiffs ‘believe [it] contains a regurgitation analysis of actual ChatGPT user conversations.’ That belief was mistaken. If you think there is a different RFP that justifies production of this request, please let us know. On Friday, I asked you to identify a relevant RFP for these requests, but you weren’t prepared to do so.*”

Maisel Decl., Ex. 20 (June 23, 2025 Email from Counsel for OpenAI). After hearing argument on June 25, 2025, this Court ordered the parties to continue meeting and conferring on the issue, “including allowing OpenAI time to investigate the datasets.” Dkt. [270](#).

C. OpenAI Obstructs the ChatGPT Output Log Sampling Process.

Based on OpenAI’s representation regarding its inability to search for News Plaintiffs’ content in its output logs, News Plaintiffs and this Court spent over eight months focused on creating a sample of ChatGPT output logs—a process that, as discussed below, OpenAI obstructed, impeded, and manipulated. First, OpenAI argued for a smaller sample based on false representations regarding its existing technical capabilities. News Plaintiffs had originally requested that OpenAI produce a 120 million sample of ChatGPT output log data covering December 2022 through November 2024. Dkt. [394](#) at 3. OpenAI objected to this sample size based,

in part, on the “technical costs,” including “more computational, data engineering, system architecture, and software engineering resources that linearly increase for each additional record.” Maisel Decl., Ex. 60 at 2 (July 18, 2025 Letter from Counsel for OpenAI). This representation is belied by Mr. Monaco’s testimony that OpenAI already had the ability [REDACTED] [REDACTED] *id.*, large datasets, such as the more than 80 million output logs. Ex. 2 at 234:5-21. Indeed, Mr. Monaco [REDACTED] [REDACTED], stating in his deposition that [REDACTED] [REDACTED] *Id.* at 23:10-23:24.

Second, once a smaller sample was carefully selected to represent, as much as possible, the entire set of ChatGPT outputs, OpenAI botched the sampling process by, *inter alia*: (1) deleting billions of ChatGPT conversations after the filing of The Times’s complaint, in violation of its discovery preservation obligations and this Court’s Preservation Order; (2) substituting millions (10%) of the logs in Plaintiffs’ requested sample because of OpenAI’s continued deletions and “processing errors” [REDACTED] [REDACTED] and (3) producing (only after this Court ordered it to do so) the 20 million sample with so many redactions as to render the sample, in the Court’s own words, “unusable.” Maisel Decl., Ex. 21 (Jan. 15, 2026, Hr’g Tr.) at 50:25-51.

1. OpenAI Deleted or Compressed Billions of ChatGPT Conversations, Thereby Rendering Them Unsearchable.

By the time The Times filed its suit in December 2023, OpenAI was required to “suspend its routine document retention/destruction policy and put in place a ‘litigation hold’ to ensure preservation of relevant documents.” *Zubulake v. UBS Warburg LLC*, 220 F.R.D. 212, 218 (S.D.N.Y. 2003). The Times reminded OpenAI of this requirement on January 23, 2024, when it

wrote to OpenAI regarding its “oblig[ation] to preserve all documents related to: . . . 4) existing or potential user requests, traffic, or behavior when interacting with the Models; 5) output generated by the Models, including output that repeats or summarizes training material or contains hallucinations or misinformation.” Maisel Decl., Ex. 22 (Jan. 23, 2024 NYT Preservation Ltr. to OpenAI) at 2; *id.*, Exhs. 61 (Feb. 29, 2024, OpenAI letter to The Times) and 62 (Mar. 5, 2024, email from The Times to OpenAI). A year later, on January 13, 2025, News Plaintiffs moved for an order requiring OpenAI to “preserve its output log data, including output log data that contains News Plaintiffs’ content, hereinafter.” *Times* Dkt. [379](#) at 3. OpenAI responded on January 16: “OpenAI does not have the preexisting technological functionality to fulfill Plaintiffs’ request to somehow preserve only user conversations that reference Plaintiffs’ content (to the extent all of Plaintiffs’ content could even be systematically identified).” *Times* Dkt. [423](#) at 2. OpenAI’s representation turned out to be false.

As News Plaintiffs previously summarized, and the parties have extensively briefed:¹⁵ “On at least 24 separate occasions since the filing of The Times’s complaint, OpenAI compressed tens of billions of consumer ChatGPT logs, which rendered them unsearchable.” Dkt. [1061](#) at 1. “OpenAI further: (a) destroyed virtually all of its API logs, (b) represented that it does not have possession, custody, or control of the enterprise ChatGPT logs; and (c) destroyed billions of consumer ChatGPT logs where the user turned ChatGPT’s preservation function off or were the subject of a user-initiated deletion.” *Id.* During the period from the launch of ChatGPT in 2022 through the week of March 24, 2025, OpenAI reported that it had deleted about [REDACTED] of all

¹⁵ The voluminous filings on OpenAI’s deletion of output log data are documented in Appendix A of the accompanying Maisel Declaration. *See also* Dkt. [1061](#), Appendix A (listing filings); Maisel Decl. Exhs. 54-56 (letter correspondence re. preservation), 57 (excerpts of Jan. 22, 2025, Hearing Transcript on preservation dispute).

conversations, including approximately [REDACTED]
[REDACTED] ChatGPT consumer conversations generated through the “Temporary Chat” feature. Maisel Decl., Ex. 23 (OpenAI’s Nov. 5, 2025 Supp. Resp. to Daily News Interrog. No. 14) at 10-11. From April through July 2025, [REDACTED] ChatGPT consumer conversations were generated through the ‘Temporary Chat’ feature, and approximately [REDACTED] ChatGPT consumer conversations were [REDACTED].” *Id.*

When News Plaintiffs learned that OpenAI was not preserving output logs, it first addressed the matter with OpenAI and then raised the issue with the Court. On May 13, 2025, this 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.” Dkt. [33](#) at 2 (emphasis in original). Even setting aside OpenAI’s preservation obligations *before* the Court’s order, as of May 13, 2025, OpenAI’s court-ordered obligation to preserve and segregate the output logs could not have been clearer. *See* Maisel Decl., Ex. 24 (Feb. 11, 2026 Hearing Tr.) at 15:9-11. (“And we do have an ongoing issue about how and whether my May 13 order was complied with, given how explicit it was.”).

OpenAI did not comply with this Court’s Order. As later became clear when OpenAI was ordered to produce the 20 million output logs, OpenAI continued to delete logs. In the short time period between the agreement and its production, OpenAI allowed over [REDACTED] of the logs selected for the 20 million sample to be deleted. *See* Maisel Decl., Ex. 65 at 7-8, 13-15. As Mr. Monaco conceded, OpenAI [REDACTED] Ex. 2 at 277:4. When asked whether OpenAI had taken any steps to comply with the Court’s order, Mr. Monaco said [REDACTED]
[REDACTED]

[REDACTED] *Id.* at 276:20-25. In other words, OpenAI thought about complying with the Court’s

order; decided complying would be hard; and thus took no steps to do so. In fact, OpenAI’s data shows [REDACTED]

[REDACTED] Sam Altman and OpenAI stridently and repeatedly denounced the Preservation Order on public websites.¹⁶ Maisel Decl., Ex. 25 (Monaco Jan. 27, 2026 30(b)(6) Dep. Tr.) at 95:3-6 [REDACTED]

[REDACTED] Ex. 26 (Monaco 5); Ex. 27 (Opening Expert Report of Dr. Tom Goldstein), ¶ 436.

Meanwhile, all along OpenAI had the capabilities, and in fact used those capabilities, to search ChatGPT output logs for News Plaintiffs’ content.

First, as part of its Giraffe “evaluations,” OpenAI developed search tools to look for [REDACTED] in the ordinary course of business to search large datasets of ChatGPT conversations [REDACTED] See Section I(B).

Second, shortly after The Times filed its Complaint, OpenAI searched its API output logs for the examples of regurgitation attached as Exhibit J to The Times’s complaint. *See* Maisel Decl., Ex. 28 (Jan. 29, 2025 Trinh 30(b)(6) Dep. Tr.) at 97:8-97:13, 104:3-104:21 [REDACTED]

[REDACTED]

[REDACTED] That is, OpenAI was willing and able to search its output logs—when it benefitted OpenAI.

¹⁶ *See* Maisel Decl., Ex. 29 (Sam Altman June 5, 2025 X Post) (“Recently the NYT asked a court to force us to not delete any user chats. we think this was an inappropriate request that sets a bad precedent. we are appealing the decision. We will fight any demand that compromises our users’ privacy; this is a core principle.”); Ex. 30 (OpenAI June 5, 2025 Blog Post titled “How we’re responding to The New York Times’ data demands in order to protect user privacy.”).

Third, as part of this Court’s Order to conduct a spoliation sampling exercise, OpenAI searched its extended 30-day tables.¹⁷ Dkt. 79. That spoliation exercise showed: [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

OpenAI failed to mention these capabilities in any of the five sworn Monaco declarations it submitted to this Court:

“I understand that the plaintiffs have asked OpenAI to perform a search over the full extant output log data for their content . . . ***OpenAI’s current infrastructure also does not include any mechanism that could search a large dataset of output logs like decompressed Conversation Data for specific content.*** OpenAI would need to design a custom system that allows using ‘n-grams’ or fuzzy searches to look for instances of Conversation Data that contain strings of words that are similar to some other specific string of words. Designing a proof-of-concept for such a system would ***take at least [REDACTED] of dedicated full-time engineer time.*** Additional work would then be required to achieve a working prototype, followed by non-trivial ongoing maintenance engineering work to ensure ongoing reliability. The system would be very costly and burdensome to build and maintain. OpenAI cannot precisely estimate the costs necessary to operate such a search mechanism without a more detailed specification. But based on information available, ***I estimate that this would cost OpenAI at least [REDACTED] up front to build and would likely require a significant amount of additional costs for each search.*** These costs would be in addition to the costs I described in the prior paragraph [REDACTED] in up front costs and at least [REDACTED] in ongoing costs].”¹⁸

¹⁷ OpenAI saves its output logs in searchable format in “30-day tables” before deleting or compressing the data for longer term storage. See Maisel Decl., Ex. 31 (May 27, 2026 Discovery Conference PM Tr.) at 36:16-19 (OpenAI explaining that OpenAI has a “30-day preservation period for even stuff that has been designated for deletion so that we can go back.”).

¹⁸ Compare Maisel Decl., Ex. 3 (V. Monaco May 23, 2025 Decl.) with Dkt. [1337](#) at 1-2 (“To be clear: the copies of ChatGPT conversations that Plaintiffs demand in their Motion [the 78 million reservoir] are not ‘new’ sources of data. To the contrary, they are drawn from the same universe of conversation data that is represented by the 20M Sample. Put another way, none of the copies of ChatGPT conversations that Plaintiffs reference in their motion contain data that was not logged in the data source that OpenAI used to derive the 20M Sample”)

“I am not aware of *any prior effort* through which OpenAI was able to pull individual records of Conversation Data from OpenAI’s primary offline storage system for analysis or any other purpose.”¹⁹

“Based on my knowledge and experience with OpenAI’s data pipelines and related cloud platform expenses, and as described further below, I realistically estimate that *it would require up to [REDACTED] and roughly [REDACTED]* in compute costs to retrieve, decompress, process, de-identify and store a 20 million record sample of Conversation Data, and *up to [REDACTED] and roughly [REDACTED]* in compute costs to retrieve, decompress, process, de-identify and store a 120 million record sample, assuming it is even possible to extract and process this many records at once.”²⁰

“OpenAI plans to use the same API to de-identify the sampled Conversation Data after it has been retrieved, decompressed, and processed. *However, even a sample of 20 million records is far larger than any dataset I have previously scrubbed using this API*²¹. . . Given this constraint, I estimate that it will take approximately three to four weeks (with one fulltime engineer working at 20% capacity) to de-identify 20 million records of Conversation Data and cost OpenAI [REDACTED] in compute costs.”²²

Mr. Monaco’s representations to this Court were false. Instead of preserving and searching ChatGPT conversations for News Plaintiffs’ content each month, OpenAI deleted or compressed billions of ChatGPT conversations, thereby rendering them inaccessible to News Plaintiffs for discovery purposes.

2. OpenAI’s Substitution of Millions of Plaintiffs’ Requested Logs Further Skewed the 20 Million Sample.

In addition to deleting ChatGPT conversations, OpenAI further skewed the 20 million sample to underrepresent conversations that invoked retrieval-augmented generation (“RAG”), a process through which OpenAI obtained News Plaintiffs’ online content in real time to respond to user queries. As discussed above, OpenAI substituted over a million conversations (5%) because

¹⁹ Maisel Decl., Ex. 32 (V. Monaco Aug. 4, 2025, Decl.) ¶ 2.

²⁰ *Id.* ¶ 3.

²¹ Mr. Monaco later testified that a far larger dataset, the 78 million, had already been [REDACTED]
[REDACTED] Ex. 2 at 245:18-23.

²² Ex. 32, ¶ 9.

of OpenAI's ongoing deletions during the pendency of this Court's Preservation Order. OpenAI further substituted another 731,402 conversations due to a "processing error." Maisel Decl., Ex. 33 (Jan. 22, 2026 Email from Counsel for OpenAI).²³ Mr. Monaco testified that a processing error occurred when the [REDACTED]

[REDACTED] Maisel Decl., Ex. 34 (V. Monaco Jan. 28, 2026 Dep. Tr.) at 22:25-23:18. News Plaintiffs' expert has concluded [REDACTED]

[REDACTED] Ex. 26, ¶442. The end result of OpenAI's deletions and substitutions is that only approximately [REDACTED] of the 20 million sample [REDACTED] *Id.*, ¶ 444.

[REDACTED] *Id.* ¶¶ 444-447. Unfortunately, only the 20 million sample—plus a [REDACTED] recently produced on May 28, 2026²⁴—contains *any* RAG conversations at all.

3. OpenAI Reneged On Its Agreement to Produce the 20 Million Logs, Then, When Ordered to Do So, Over-Redacted Them So as to Render Them "Unusable."

After News Plaintiffs agreed to accept OpenAI's own 20 million log proposal as an initial step so they could begin their review, while reserving rights to seek more logs that would constitute a true representative sample, OpenAI changed course and refused to produce the 20 million

²³ OpenAI also substituted ChatGPT conversations that belonged to an [REDACTED]

[REDACTED] Ex. 33.

²⁴ News Plaintiffs are still analyzing [REDACTED], and intend to submit a supplemental expert report.

conversation sample. Dkt. [656](#) at 1; Maisel Decl., Exhs. 63-64.²⁵ The Court then ordered production, directing OpenAI “to produce the 20 million de-identified Consumer ChatGPT Logs to News Plaintiffs by November 14, 2025, or within 7 days of completing the de-identification process.” Dkt. [734](#) at 2. OpenAI then sought reconsideration from this Court, while simultaneously filing an objection to Judge Stein. In denying OpenAI’s motion for reconsideration, this Court concluded with this important observation:

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

Dkt. [896](#) at 9, n.9. Judge Stein affirmed Judge Wang’s Order on January 5, 2026. Dkt. [1021](#).

On December 15, 2025, OpenAI finally produced this sample with only two months left in fact discovery—but used an AI tool to apply over **19 billion** redactions. Maisel Decl., Ex. 35 (Decl. of Krithika Rao at Dkt. [1061-5](#)), ¶ 3; *id.*, Ex. 66 (OpenAI describing its redactions as “intentionally robust”). The materials had been so over-redacted that this Court described it as “mind boggling.” Ex. 20 at 50:25-51. As this Court asked OpenAI: “Maybe you can explain to me how there are 19 billion redactions over I don’t know how many millions of logs it comes out to, something like over a thousand redactions per unit, 15.4 million logs. How does that work?” *Id.* at 40:23-41:5. OpenAI answered that it had used an AI redaction tool, which OpenAI had “augmented” in some unspecified way—that is, a tool that OpenAI did not use in the ordinary course of its business. *Id.* As this Court noted, OpenAI “took that [AI] tool and [OpenAI] added additional layers of

²⁵ News Plaintiffs initially requested a much larger sample, which OpenAI refused. Maisel Decl., Exhs. 58-59 (letter correspondence re. ChatGPT output log sampling).

redactions to the point that the data is unusable. And that is not an excuse after you’ve been fighting for months about providing these logs.” *Id.* at 48.

OpenAI did not produce the 20 million-log “sample” with fewer redactions until January 22, 2026—a week before the end of fact discovery. Even then, a large number of redactions remain, including to News Plaintiffs’ domains, names, and other fields, which has hampered News Plaintiffs’ searches over the data. Ex. 26, ¶¶ 448-451.

The entire time that OpenAI was engaging in the improper over-redaction of this sample, it had in its possession a sample of 78 million conversations that had already been de-identified.

D. OpenAI Obstructs Mr. Monaco’s First Deposition.

On January 27-28, 2026, News Plaintiffs deposed Mr. Monaco, OpenAI’s 30(b)(6) witness on its investigations into regurgitated copyrighted content. Notwithstanding Mr. Monaco’s lack of preparation and OpenAI’s interference with the deposition (facts that this Court found in Dkt. [1492](#)), Mr. Monaco’s testimony revealed three important points: [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Three days later, News Plaintiffs moved to compel production of OpenAI’s Project Giraffe evaluations and the 78 million ChatGPT conversations. Dkt. [1220](#). OpenAI continued to misrepresent its search capabilities and vehemently resisted production:

- OpenAI claimed it could not search “*all* its outputs for any training data.” Dkt. [1267](#) at 1 (emphasis added).
- OpenAI said it did not produce the ChatGPT dataset because, *inter alia*, “in order to provide the data out to a party outside the company, there’s a de-identif[ication] process that everybody is talking about and there is the objection on user privacy grounds . . .” Ex. 23 at 21:13-16; *see also* Dkt. [1295](#) (OpenAI Feb. 12, 2026 Letter to Court) at 2 [REDACTED]

- [REDACTED]
- Regarding the Project Giraffe evaluations, OpenAI told the Court: “This is not an analysis that is trying to answer, I think the precise questions the plaintiffs are trying to answer. This is an analysis trying to answer a different question, although I recognize it’s closely related to the questions the plaintiffs really want answered” Ex. 23 at 21:19-24.
 - OpenAI represented that it had not searched the 78 million output logs [REDACTED]. Dkt. [1295](#) at 1.
 - OpenAI: “It makes no sense for Plaintiffs to now demand—less than two weeks before the close of fact discovery—that OpenAI undertake the massively burdensome process of producing non-representative, incomplete sources of user data that are both irrelevant and useless to this case.” Dkt. [1337](#) at 1.

E. Mr. Monaco’s Second Deposition Reveals OpenAI’s Search Capabilities and Already-Conducted Searches for News Plaintiffs’ Content.

When Mr. Monaco testified as OpenAI’s Rule 30(b)(6) witness for a second time on April 8, 2026, he made a number of stunning admissions, including that OpenAI had the ability to search output logs, compiled datasets of 78 million and 10 million output logs (collectively referred to as “over 80 million output logs”) that had been de-identified, [REDACTED]. His sworn deposition contradicted representations OpenAI repeatedly made to this Court and News Plaintiffs, as set forth, *supra*, p. 1, in the introduction to this Motion.

The full facts are these: In May 2022, before the launch of ChatGPT, OpenAI began

[REDACTED]
[REDACTED] Mr. Monaco testified: [REDACTED]
[REDACTED]
[REDACTED] Ex. 2 at 212:2-5. OpenAI’s solution was to build a [REDACTED] filter to
[REDACTED]

[REDACTED] (*Times* Dkt. 1, ¶ 54; Maisel Decl., Exhs. 44-45), the

[REDACTED],²⁶ created a filter that could be used to block the regurgitation of copyrighted content.

Ex. 2 at 228:8-228:19. Specifically, after OpenAI was sued, [REDACTED]

[REDACTED]. Ex. 2 at 231:1-6: ([REDACTED])

[REDACTED] OpenAI

[REDACTED]. *Id.* at 228:4-231:6.

After ChatGPT launched, Project Giraffe staff amassed datasets of tens of millions of ChatGPT conversations to [REDACTED]. *Id.* at 243:8-13. These Project Giraffe evaluations required OpenAI to create three specialized sets of tools.

First, OpenAI built tools to search its training datasets for Plaintiffs’ content. As Mr. Monaco testified regarding [REDACTED]

[REDACTED] *Id.* at 223. OpenAI would then “prune” these datasets by leaving in place Times content for which there was a “likelihood of regurgitation” and removing content with a “regurgitation likelihood” that fell below a certain threshold. *Id.* at 216:17-217:21; Maisel Decl., Exhs. 51-52 [REDACTED]).

²⁶ Ex. 2 at 224:13-16. *See also id.* at 224:22-225:18; 225:22-226:5; 227:3-10 (identifying counsel).

Second, OpenAI built tools to de-identify and scrub personally identifying information (“PII”) from ChatGPT conversation data to [REDACTED]

[REDACTED] Ex. 2 at 249:18-22. That is, by June 2023, OpenAI had already amassed a dataset of 78 million [REDACTED] ChatGPT conversations. *Id.* at 245:18-246:3.

Third, OpenAI built tools to run tests, or “evaluations,” to assess [REDACTED]

[REDACTED]

[REDACTED] *Id.* at 241. OpenAI [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] *Id.* at 241-42. In other words, OpenAI [REDACTED]

[REDACTED]. OpenAI called

these [REDACTED] evaluations.

OpenAI [REDACTED]

[REDACTED]—referred to as the [REDACTED]

[REDACTED]—in May-June 2023, shortly after The Times sent a letter raising concerns about OpenAI’s unauthorized use of its content. *Id.* at 234; Maisel Decl. Exhs. 46-51.

In 2024, OpenAI performed [REDACTED]

[REDACTED] *Id.* at

235:25-237:25 [REDACTED]

[REDACTED] Maisel Decl., Ex. 36 (OPCO_MDL_004898640-OPCO_MDL_004898659) (evaluation metadata files).

Thus, at the time News Plaintiffs filed their lawsuits, OpenAI already possessed the tools to identify News Plaintiffs’ copyrighted materials used for LLM training; produce datasets of

searchable, de-identified, and PII-scrubbed ChatGPT conversations; and search for [REDACTED]

F. OpenAI Claims that it Already “Produced” the ChatGPT Dataset it Had Been Refusing to Produce for Months.

On May 14, 2026, News Plaintiffs asked OpenAI whether a dataset denominated [REDACTED] [REDACTED] in the evaluations was the same dataset of 78 million ChatGPT logs that this Court ordered OpenAI to produce following extensive motion practice. Maisel Decl., Ex. 37 at 1-2 (May 14, 2026 K. Logan Email). On May 18, 2026, OpenAI responded that this was the very ChatGPT dataset Plaintiffs had sought, and—despite OpenAI vehemently opposing its production²⁷—OpenAI now said that it “has been available for inspection on the training data VM for over a year (since April 2025).” *Id.* at 1 (May 18, 2026 Email from OpenAI). This makes no sense. Either OpenAI unintentionally produced the dataset and it was so hidden in the training inspection data that even OpenAI did not realize it, or OpenAI knew it buried the dataset in a previous production, but hid that fact from the Court and News Plaintiffs for nearly two years—all the while vigorously arguing that turning over these logs would violate user privacy.²⁸ This is yet another example of OpenAI’s obfuscation making the discovery process as burdensome as possible.

II. LEGAL FRAMEWORK

District courts retain wide discretion to impose sanctions for discovery misconduct under the Federal Rules of Civil Procedure and their inherent authority. *See Reilly v. Natwest Mkts. Grp. Inc.*, 181 F.3d 253, 267 (2d Cir. 1999). Under Rule 37, courts may order sanctions if a party fails

²⁷ Dkt. 1295 (OpenAI Feb. 12, 2026 Letter to Court) at 2 [REDACTED]

²⁸ The [REDACTED] dataset was produced among [REDACTED] files on the training data virtual machine.

“to obey an order to provide or permit discovery,” 37(b)(2)(A); “to provide information or identify a witness as required by Rule 26(a) or (e),” 37(c)(1); or to respond to “interrogatories under Rule 33 or a request for inspection under Rule 34,” 37(d)(1)(A)(ii). Additionally, if electronically stored information that should have been preserved is lost, upon a finding of prejudice, courts may “order measures no greater than necessary to cure the prejudice.” 37(e)(1). If the court finds a party acted with an “intent to deprive another party the information’s use in litigation,” courts may “presume the lost information was unfavorable to the party” or “instruct the jury that it may or must presume the information was unfavorable to the party.” 37(e)(2).

Rule 37 provides a menu of sanctions courts may impose, including, (i) “directing that the matters embraced in the order or other designated facts be taken as established for purposes of the action, as the prevailing party claims” and “(ii) prohibiting the disobedient party from supporting or opposing designated claims or defenses, or from introducing matters in evidence; striking pleadings in whole or in part.” Fed. R. Civ. P. 37(b)(2)(A)(i), (ii). In determining whether preclusion of evidence is merited, courts consider: (i) the party’s explanation for failing to disclose the evidence; (ii) the evidence’s importance; (iii) the prejudice to the opposing party; and (iv) whether a continuance is possible. *Patterson v. Balsamico*, 440 F.3d 104, 117 (2d Cir. 2006); *Softel, Inc. v. Dragon Medic. & Sci. Commc’ns, Inc.*, 118 F.3d 955, 961 (2d Cir. 1997).

Rule 37 “places the burden on the disobedient party to avoid” a sanction and permits a court “to consider the full record in the case in order to select the appropriate sanction.” *J.C. v. Zimmerman*, 150 F.4th 136, 146 (2d Cir. 2025) (cleaned up). When imposing sanctions under this Rule, courts in the Second Circuit analyze four non-exhaustive factors: “(1) the willfulness of the non-compliant party or the reason for noncompliance; (2) the efficacy of lesser sanctions; (3) the duration of the period of noncompliance[;] and (4) whether the non-compliant party had been

warned of the consequences of noncompliance.” *Agiwal v. Mid Island Mortg. Corp.*, 555 F.3d 298, 302 (2d Cir. 2009).

Finally, a court has inherent authority to impose sanctions, even in circumstances that would not fit within any other rule. *Chambers v. NASCO, Inc.*, 501 U.S. 32, 49 (1991) (“[T]he inherent power of a court can be invoked even if procedural rules exist which sanction the same conduct.”). Thus, a “federal court may exercise its inherent power to sanction a party or an attorney who has acted in bad faith, vexatiously, wantonly, or for oppressive reasons.” *Rossbach v. Montefiore Med. Ctr.*, 81 F.4th 124, 141 (2d Cir. 2023) (cleaned up). “This authority stems from the very nature of the courts and their need to be able to manage their own affairs so as to achieve the orderly and expeditious disposition of cases.” *Hinds v. Cnty. of Westchester*, No. 11-cv-7265, 2020 WL 7046843, at *5 (S.D.N.Y. Dec. 1, 2020) (quotations omitted); see also *Kassenoff v. Kassenoff*, No. 22-cv-2162, 2024 WL 291225, at *2 (S.D.N.Y. Jan. 25, 2024).

III. ARGUMENT

News Plaintiffs recognize that seeking sanctions against a litigation adversary should not be taken lightly. Here, this significant step is warranted given the severity of OpenAI’s discovery misconduct and non-compliance, which includes: (i) its false claims that it could not run searches over training and output log data that had already been searched; (ii) OpenAI’s continued deletion of output logs in violation of this Court’s Order and its preservation obligations; and (iii) OpenAI’s years-long failure to comply with numerous discovery requests. See *Shcherbakovskiy v. Da Capo Al Fine, Ltd.*, 490 F.3d 130, 140 (2d Cir. 2007) (“Rule 37 permits the imposition of ‘just’ sanctions; the severity of sanction must be commensurate with the non-compliance.”).

A. OpenAI Should be Precluded From Relying on the 20 Million Sample Pursuant to Rule 37(b) and (e)(1) and this Court’s Inherent Authority.

1. Preclusion Is an Appropriate Remedy Under Rule 37(b) Due to OpenAI’s Violation of This Court’s Orders.

Under Rule 37(b), preclusion of the 20 million sample is warranted in response to OpenAI’s failure to comply with this Court’s May 13, 2025 Order, which required OpenAI to preserve and segregate output logs. *See* Ex. 23 at 15. The Order was clear and explicit, yet OpenAI expunged billions of logs. OpenAI further violated the Court’s order by deleting 5% of the output logs called for by the sampling protocol and unilaterally substituting another 5% of the output logs, thereby inserting over two million logs of its own choosing. Finally, when OpenAI eventually “produced” the 20 million sample as the Court had ordered, it used an augmented AI redaction tool to apply 19 billion redactions that rendered the production “unusable.” Ex. 20 at 49 (Jan. 15, 2026).

Each of the *Agiwal* factors favors preclusion. 555 F.3d at 302. *First*, there can be no question as to the willfulness of OpenAI’s conduct, nor any excuse for its non-compliance. According to Mr. Monaco, OpenAI thought about complying with the Court’s Preservation Order, but then decided not to. *See* Ex. 2 at 276:20-277:4 [REDACTED]

[REDACTED] OpenAI then flouted Judge Wang’s order to produce the logs by creating its own redaction system that rendered the production, as this Court found, “unusable.” Ex. 20 at 49 (Jan. 15, 2026). Ultimately, OpenAI only produced the logs in response to a third order, and then only at the very end of the discovery period.

Second, lesser sanctions would not be effective. As discussed above, OpenAI had effectively “run out the discovery clock” by withholding production of the 20 million output logs

until the very end of fact discovery. Costs and fees will not suffice in themselves given the scope and duration of OpenAI's discovery misconduct. The sanction should be directed to the utility of the evidence that OpenAI's actions kept from the record for two years. That is, it should address the fact that the productions were unusable (indeed, entirely unavailable) to News Plaintiffs during the discovery period or until so late in the period as to be effectively the same thing.

Third, the duration of the period of noncompliance in this case lasted two years, from the beginning of the discovery period to the end.

Finally, there should have been no need to warn OpenAI of the consequences of its actions. OpenAI is an incredibly sophisticated company valued at one trillion dollars with no shortage of legal and technical resources. Moreover, this Court has repeatedly alerted OpenAI that its conduct potentially implicated the imposition of sanctions.

This remedy is also supported by an analysis of the preclusion factors under *Patterson*, 440 F.3d at 117. First, OpenAI can offer no reasonable explanation for continuing to delete output logs in violation of this Court's order, the unilateral substitutions it made to the 20 million sample, or the unreasonable redactions it applied. Indeed, its failure to disclose is not limited to one category of information but instead reflects a deliberate and systemic effort to obstruct discovery into OpenAI's existing output logs and analysis. *See, e.g.*, Dkt. [1419](#) (order finding "that the pattern of objections and Mr. Monaco's resulting answers impeded, delayed and frustrated the fair examination of OpenAI, the deponent."); Dkt. [1492](#) ("[I]t is not clear whether Mr. Monaco's insufficient testimony and nonresponsive answers were due to insufficient preparation, improper preparation, a misapprehension of the requirements of Rule 30(b)(6), or a combination of these factors"). To the degree OpenAI tries to justify its noncompliance based on burden and privacy concerns, OpenAI filed motions and appeals making this argument, which both Magistrate Judge

Wang and Judge Stein rejected.²⁹ Second, as described above (Section I(C)), OpenAI’s conduct distorted the discovery process, leading News Plaintiffs and this Court down the garden path to the 20 million sample and then distorting that sample through its deletions. As a result, OpenAI impeded the very thing this process was intended to produce: a representative sample. Ex. 26, ¶¶ 437, 447. Third, for the same reason, the only prejudice to OpenAI in precluding the 20 million sample is that it cannot benefit from its manipulation of the sample. Finally, OpenAI has represented that its deletions cannot be retrieved or restored. A continuance would therefore not resolve the prejudice suffered.

Serious sanctions are especially appropriate here because OpenAI’s conduct, including the violation of the Court’s preservation Order, was knowing and intentional. *See Brown v. City of New York*, No. 15-cv-4488, 2018 WL 3193208, at *2 (E.D.N.Y. June 28, 2018) (upholding magistrate judge’s order issuing preclusion and monetary sanctions where defendants negligently failed to produce a key audio recording until after discovery closed, blaming the delay on “inadvertent clerical error”); *New York State Nat. Org. for Women v. Cuomo*, 1997 WL 671610, at *5 (S.D.N.Y. Oct. 28, 1997) (preclusion sanctions “may be imposed where the failure to meet discovery obligations is not willful, but merely negligent”). Beyond the egregious redactions, OpenAI previously [REDACTED]

but deleted content called for in discovery. Pursuant to Rule 37(b)(2)(A), this Court should

²⁹ In denying OpenAI’s Rule 72(a) objections and affirming Magistrate Judge Wang’s Order directing OpenAI to produce the 20 million sample, Judge Stein held OpenAI’s reliance on *S.E.C. v. Rajaratnam*, 622 F.3d 159 (2d Cir. 2010), a case addressing the legality of the wiretaps at issue, was misplaced. Dkt. [1021](#) at 3. Compared to *Rajaratnam*, the Court held that “the privacy interests in users’ conversations with ChatGPT which users voluntarily disclosed to OpenAI and which OpenAI retains in the normal course of its business” was weaker. *Id.* The Court additionally noted that “OpenAI identifies no caselaw requiring a court to order the least burdensome discovery possible.” *Id.*

preclude OpenAI from relying on the 20 million sample, thereby benefiting from its misconduct. Fed. R. Civ. P. 37(b)(2)(A)(ii) (“prohibiting the disobedient party from . . . introducing designated matters in evidence.”). By contrast, News Plaintiffs should be permitted to introduce the 20 million sample, not as a representative data set, but for the limited purpose of showing, for example, how OpenAI’s products and models operate and use News Plaintiffs’ content. Such limited use is necessary as OpenAI informed News Plaintiffs it has no technical documentation in connection with its output logs. *See* Maisel Decl., Ex. 17 (Mar. 19, 2025 Letter) at 2-3 (stating OpenAI does not, *inter alia*, “have documentation detailing the meaning of all the fields in conversation logs,” or documentation about classifiers).

This Court has precluded evidence in analogous instances of noncompliance with court orders. *See, e.g., Joint Stock Co. Channel One Russia Worldwide v. Infomir LLC*, No. 16-cv-1318 (GBD), 2019 WL 4727537, at *1 (S.D.N.Y. Sept. 26, 2019), *aff’d sub nom. Joint Stock Co. “Channel One Russia Worldwide” v. Infomir LLC*, No. 16-cv-1318 (GBD), 2020 WL 1479018 (S.D.N.Y. Mar. 26, 2020). In *Channel One Russia Worldwide*, Russian television broadcasters alleged the unauthorized interception and retransmission of plaintiffs’ programming to U.S. consumers. Plaintiffs relied on an affidavit from a paralegal, who claimed that a set-top box connecting to a domain associated with the defendants received an unauthorized Channel One stream. *Id.* at *4. Plaintiffs failed to timely produce the critical file, referred to as the Channel One PCAP, for more than two years, despite mandatory disclosure obligations, document requests, a subpoena, and multiple court orders requiring production of documents or ESI reflecting the investigation. *Id.* at *1. The Court granted defendants’ motion for sanctions and ordered that the plaintiffs were precluded from relying on, *inter alia*, the paralegal affidavit and the Channel One PCAP. *Id.* at *20.

A similar remedy is appropriate here. Preclusion of the 20 million sample is appropriate where OpenAI “ultimately produced only after violating a series of court orders and facing spoliation sanctions.” *Id.* at *31. Indeed, OpenAI’s conduct is more egregious than that before the Court in *Channel One Russia Worldwide*, where the information was ultimately disclosed. OpenAI deleted or replaced 10% of the 20 million sample, and they cannot be retrieved. In these circumstances, News Plaintiffs’ request to preclude OpenAI from relying on the 20 million sample “is well-tailored to ensure that [OpenAI] will not benefit from their own failure to comply with this Court’s order.” *Id.* (citation omitted).

2. Preclusion Is an Appropriate Remedy Under Rule 37(e)(1) Due to OpenAI’s Failure to Preserve Output Logs.

This remedy is also appropriate under Rule 37(e)(1) due to OpenAI’s deletion of output logs from the 20 million sample, “information that should have been preserved.” *Medcenter Holdings Inc. v. Web MD Health Corp.*, 734 F. Supp. 3d 303, 313 (S.D.N.Y. 2024) (precluding publisher from offering “any evidence relying on or derived from the data” lost from the database under Rule 37(e)(1)). OpenAI had an obligation to preserve output logs both before and after this Court’s May 2025 Preservation Order. *Fujitsu Ltd. v. Fed. Express Corp.*, 247 F.3d 423, 436 (2d Cir. 2001) (“The obligation to preserve evidence arises when the party has notice that the evidence is relevant to litigation or when a party should have known that the evidence may be relevant to future litigation.”). Since December 2023, when The Times filed its initial Complaint, OpenAI was on notice of its preservation obligations in connection with log data. This obligation was emphasized in correspondence between the parties, at hearings, and in meet and conferrals. *See, e.g.*, Ex. 22 (Times Jan. 23, 2024 Letter to OpenAI reiterating OpenAI’s obligation “to preserve all documents” related to, *inter alia*, “output generated by the Models, including output that repeats or summarizes training material or contains hallucinations or misinformation.”).

In violation of this Court’s orders and its preservation obligations, OpenAI has delayed production of, deleted, or otherwise rendered unavailable vast quantities of relevant information, including compressing [REDACTED] of consumer ChatGPT logs, destroying all API logs, and deleting [REDACTED] output logs that had been designated by News Plaintiffs through the sampling exercise. OpenAI deleted output logs “while it knew or should have known of its duty to preserve this evidence.” *Charlestown Cap. Advisors, LLC v. Acero Junction, Inc.*, 337 F.R.D. 47, 67 (S.D.N.Y. 2020); *see also Skyline Steel, LLC v. PilePro, LLC*, 101 F. Supp. 3d 394, 408 (S.D.N.Y. 2015) (cleaned up) (“If triggered, the preservation obligation requires a litigant to do more than refrain from intentionally destroying relevant evidence; the litigant must also take affirmative steps to prevent inadvertent spoliation”).

OpenAI’s deletion of output logs, including those which should have been preserved as part of the 20 million sample, has prejudiced News Plaintiffs. Far from an ancillary dataset, output logs are central to several of News Plaintiffs’ claims. The importance of this data has been recognized repeatedly by the Court and OpenAI. *See e.g., Times* Dkt. [72](#) at ¶ 41 (“Plaintiff’s complaint focuses heavily on the ‘outputs of Defendants’ GenAI models.’”). Accessing output data from OpenAI products that contain infringing reproductions of News Plaintiffs’ copyrighted works is important to News Plaintiffs’ output-based claims. *See, e.g., Arista Records, Inc. v. Mp3Board, Inc.*, No. 00-cv-4660, 2002 WL 1997918, at *3 (S.D.N.Y. Aug. 29, 2002) (Stein, J.). Although News Plaintiffs need not establish that the deleted output logs contained a “smoking gun,” from the very beginning of these cases, all parties were aware that output logs constituted some of the most important evidence in this case. That alone—even without the other misconduct involved in the spoliation—would be sufficient to warrant severe sanctions. *Karsch v. Blink Health Ltd.*, 17-cv-3880 (BCM), 2019 WL 2708125, at *21 (S.D.N.Y. June 20, 2019) (quotations omitted)

(holding the moving party need not establish the loss of a “smoking gun” and that “[i]t is sufficient if the existing evidence plausibly suggests that the spoliated ESI could support the moving party’s case.”); *see also Doubleline Capital LP v. Odebrecht Fin., Ltd.*, No. 17-cv-4576, 2021 WL 1191527, at *7 (S.D.N.Y. Mar. 30, 2021) (finding prejudice where evidence “would be helpful to [plaintiffs] in connection with establishing several elements of . . . [their] claims.”).

OpenAI’s deletions, substitutions, and redactions impaired News Plaintiffs’ ability to ascertain the scope and degree to which OpenAI scrapes, grounds, and outputs News Plaintiffs’ content. In fact, approximately [REDACTED] of the skewed 20 million ChatGPT conversations sample is presently the principal source of evidence for showing OpenAI’s infringement based on its RAG capabilities because the 78 and 10 million reservoirs of ChatGPT conversations do not have RAG data. Preclusion of the 20 million is an appropriate remedy for this misconduct.

B. This Court Should Find That the ChatGPT Output Logs Include (Or Would Have Shown if Properly Produced in Discovery) Substantial and Systematic Grounding on and Regurgitation of News Plaintiffs’ Copyrighted Material Into User Outputs, and OpenAI Should Be Prohibited From Arguing to the Contrary.

Due to OpenAI’s misconduct throughout the discovery process, in addition to what the evidence has already established in this case, News Plaintiffs also seek an order finding that ChatGPT’s output logs include substantial and systematic grounding on and regurgitation of News Plaintiffs’ copyrighted material into user outputs. Under each of Rule 37(b), (c), (d), and (e)(1), courts may “direct[] that . . . designated facts be taken as established for purposes of the action, as the prevailing party claims” and/or “prohibit[] the disobedient party from supporting or opposing designated claims or defenses,” Fed. R. Civ. P. 37(b)(2)(A)(i), (ii). The Court should do so here.

During fact discovery, OpenAI knew that it had information directly responsive to multiple requests for production and refused to produce it, or in some cases, denied its existence. For

example, one of News Plaintiffs’ requests for production called for “[q]uery, session, and chat logs reflecting or analyzing user sessions related to Times Content.” Maisel Decl., Ex. 13 at 13 (RFP No. 45); *see also id.* at 10 (RFP No. 39). Another seeks: “[d]ocuments related to Defendants’ ability to identify and monitor user sessions where Defendants’ Generative AI Products and Services output copyrighted content (or suspected copyrighted content) to the user.” Maisel Decl., Ex. 38 (The Times Nov. 19, 2024 Fourth Set of Request for Production to OpenAI) at 10 (RFP No. 119); *see also id.* at 11 (RFP Nos. 120-122).³⁰ As noted—but bears repeating—Mr. Monaco testified on behalf of OpenAI that, [REDACTED]

[REDACTED] See Section I(E). Yet OpenAI failed to provide that critical information during the two-year discovery period.

OpenAI did not just oppose production of this evidence based on burden or relevance; it falsely represented to the Court that obtaining this evidence was beyond its capabilities without the expenditure of [REDACTED] and months of work and that it would be just as easy for Plaintiffs to do this work—without disclosing that this work had already been done. OpenAI’s representation that it could not search [REDACTED] *see* Ex. 17 at 1, while withholding it had the capability to search over [REDACTED] conversations, does nothing to absolve its noncompliance. The Federal Rules do not permit a responding party to evade otherwise discoverable information through hyper-technical interpretations or semantic gamesmanship. *Polaroid Corp. v. Casselman*, 213 F. Supp. 379, 381 (S.D.N.Y. 1962) (“The ends of justice are served, not by giving one side a vested right to exhaust the other, but by affording

³⁰ The Daily News Plaintiffs served similar discovery on June 24, 2024 and November 20, 2024. Maisel Decl., Exhs. 7 and 39.

both an equal opportunity to a full and fair adjudication on the merits”); Fed. R. Civ. P. 37 advisory committee’s note to 1993 amendment (“Interrogatories and requests for production should not be read or interpreted in an artificially restrictive or hypertechnical manner to avoid disclosure of information fairly covered by the discovery request, and to do so is subject to appropriate sanctions”). Under Rule 26(e), OpenAI was required to supplement its responses to requests for information targeting regurgitation datasets and evaluations. Having failed to do so, this Court is empowered to impose any of the 37(b)(2)(A) remedies, which include embracing designated facts as established, 37(b)(2)(A)(i), and/or preventing a disobedient party from introducing designated matters in evidence, 37(b)(2)(A)(ii). Fed. R. Civ. P. 37(c)(1)(C).

This sanction is also appropriate due to OpenAI’s failure to produce the same information in response to multiple interrogatories and requests for inspection. *See* Fed. R. Civ. P. 37(d)(1)(A), (3). For example, Times Interrogatory No. 9 requested that OpenAI, “Identify the existence, custodian, location, and general description of documents reflecting user sessions . . . including any logs.” Maisel Decl., Ex. 67 (Times First Set of Interrogatories to OpenAI) at 6. Similarly, Times Request for Inspection No. 6 sought, “Query, session, and chat logs related to Times Content reflecting user sessions on Your Generative AI Products and Services (‘Session Data’), including, for each session, user queries paired with responses to those queries.” Ex. 15 at 2. None of the information described above was produced in response to these requests. OpenAI withheld responsive and critical repositories of information until after the close of fact discovery, including the recently produced additional Giraffe evaluations, which were only revealed in response to News Plaintiffs’ questioning during Mr. Monaco’s second deposition.

This remedy is also merited under Rule 37(b) in response to OpenAI’s repeated violations of this Court’s meet-and-confer orders. As News Plaintiffs later learned, after two years of meet-

and-confer exchanges, multiple hearings, and several “settlement conferences” in Judge Wang’s courtroom, OpenAI violated this Court’s orders by failing to discuss a path to production in good faith—up to and including its withholding of the critical information that it had [REDACTED]

[REDACTED]

[REDACTED]

OpenAI’s conduct cannot be credibly justified as a good faith error. *In re Keurig Green Mountain Single-Serve Coffee Antitrust Litig.*, 341 F.R.D. 474, 493 (S.D.N.Y. 2022) (“the culpable state of mind for purposes of spoliation sanctions under Rule 37(b) requires a finding that the party acted knowingly in bad faith, through gross negligence, or through ordinary negligence”); *see also Oz v. Lorowitz*, No. 09-cv-5532, 2011 WL 803077, at *2 (S.D.N.Y. Mar. 7, 2011) (cleaned up) (Rule 37(b) applies “notwithstanding a lack of willfulness or bad faith,” although such factors may be “relevant to the sanction to be imposed for the failure.”). The same OpenAI employee involved in providing five sworn declarations to this Court was the individual involved in creating the [REDACTED] Further, that OpenAI employee confirmed [REDACTED] See Maisel Decl., Ex. 40 (OpenAI’s Aug. 27, 2025 MDL Privilege Log, Entry No. 42) (noting the [REDACTED] was “prepared at the direction of” current outside counsel of record). Despite this knowledge, OpenAI repeatedly represented to the Court and the parties that building a tool that could search and preserve Plaintiffs’ content would be essentially impossible. On March 31, 2025, OpenAI wrote: “To the extent OpenAI could even create a tool that could identify and preserve a subset of conversations based on as-yet-undefined criteria, doing so would require significant and highly burdensome investment of time and resources and—if even possible—would take at least several months of development.” Maisel Decl., Ex. 41 (Mar. 31, 2025 Letter from Counsel for OpenAI)

at 1. This Court also found OpenAI improperly obstructed Mr. Monaco’s first deposition. Dkt. [1419](#).

Over countless meet and confer exchanges, calls, and hearings—including conferences in this Court—not once did OpenAI tell News Plaintiffs that it already had searchable, de-identified, and PII-scrubbed reservoirs of tens of millions of ChatGPT conversations. Instead, as characteristic of the meet-and-confer process in this case, when presented with direct questions about the ChatGPT dataset—including whether it was “analyzed to identify instances of infringement” or was “part of a regurgitation-related pipeline,”—OpenAI responded that News Plaintiffs’ “belief was mistaken.” Ex. 20. OpenAI’s violation of this Court’s orders to meet and confer in good faith delayed and obstructed the production of relevant information during the discovery period.

OpenAI’s partial and delayed discovery productions do not shield it from sanctions. As the Second Circuit recognized, “hopelessly belated compliance should not be accorded great weight. Any other conclusion would encourage dilatory tactics, and compliance with discovery orders would come only when the backs of counsel and the litigants were against the wall.” *Cine Forty-Second St. Theatre Corp. v. Allied Artists Pictures Corp.*, 602 F.2d 1062, 1068 (2d Cir. 1979); *see also Fendi Adele v. Filene’s Basement, Inc.*, No. 06-cv-244 (RMB), 2009 WL 855955, at *5 (S.D.N.Y. Mar. 24, 2009) (“even if the discovering party ultimately obtains, through Herculean and expensive effort, the discovery to which it was entitled, that does not preclude it from being awarded appropriate relief to remedy the injury”).

Here, the over 80 million output logs and other Project Giraffe-related discovery were only discovered through taxing, time-consuming investigative work by News Plaintiffs and then only

produced after Court intervention, after the close of fact discovery, or both. OpenAI's years-long pattern of producing discovery only after obfuscation and delay should not be condoned.

C. This Court May Also Impose Each of the Requested Remedies Pursuant to Rule 37(e)(2) and Should Instruct the Jury That OpenAI Deleted Billions of Output Logs.

This evidence also establishes that OpenAI “acted with the intent to deprive [News Plaintiffs] of the information’s use in the litigation.” Fed. R. Civ. P. 37(e)(2); see *Hoffer v. Tellone*, 128 F.4th 433, 439 (2d Cir. 2025) (holding an intent to deprive must only be established by a preponderance of the evidence). Accordingly, this Court may alternatively award the requested remedies pursuant to Rule 37(e)(2) and should issue an instruction to the jury that OpenAI deleted billions of output logs.

As discussed in Section I(C)(i) above, at the beginning of this case News Plaintiffs asked this Court to order OpenAI to search each month’s output logs for News Plaintiffs’ content before deleting or compressing those output logs. OpenAI opposed this request, arguing to the Court that it did not have the technical ability to search the output logs for News Plaintiffs’ content. The Court and News Plaintiffs now know that representation was false, as evidenced by OpenAI’s Giraffe evaluations, its search of API output logs, and the spoliation sampling exercise—each of which required the searching of output logs for specific content. What is more, OpenAI misled the Court and News Plaintiffs regarding its existing technical capacity all while continuing to compress and delete billions of conversation logs.

OpenAI’s misrepresentations and ongoing deletions underscore its intent. *CBF Industria de Gusa S/A v. AMCI Holdings, Inc.*, No. 13-cv-2581, 2021 WL 4190628, at *18 (S.D.N.Y. Aug. 18, 2021) (quotation omitted) (intent can be inferred when the data loss cannot be “credibly explained other than bad faith”); see also *Fashion Exch. LLC v. Hybrid Promotions, LLC*, No. 14-

cv-1254, 2021 WL 1172265, at *6 (S.D.N.Y. Mar. 29, 2021) (Stein, J.) (“An intent to deprive the opposing party of the use of documents may be inferred from the party’s failure to preserve documents, even when the party that lost the documents established a credible explanation of destruction.”).

In sum, OpenAI knew that the kinds of dataset and search tools used to create Project Giraffe existed at the same time it intentionally led this Court and News Plaintiffs to believe otherwise. Where, as here, a party had direct knowledge of the technological capabilities being denied, the intent to deprive is manifest.

D. This Court Should Award News Plaintiffs Attorneys’ Fees in Connection with Their Motions and Meet and Confers Regarding Project Giraffe, Depositions of Mr. Monaco, and Time Their Lawyers and Experts Spent Reviewing Training and Output Material in the “Sandbox.”

In addition to the remedies requested above, News Plaintiffs are entitled to an award of attorneys’ fees and costs incurred as a direct result of OpenAI’s discovery misconduct. Under Rule 37(b)(2)(C), when a court grants a motion for sanctions, the court must, after giving an opportunity to be heard, require “the disobedient party, the attorney advising that party, or both to pay the reasonable expenses, including attorney’s fees, caused by the failure, unless the failure was substantially justified or other circumstances make an award of expenses unjust.” Similarly, under Rule 37(d)(3), sanctions “must” include reasonable expenses and attorneys’ fees unless the failure was “substantially justified.” *Novak v. Wolpoff & Abramson LLP*, 536 F.3d 175, 178 (2d Cir. 2008). This provision “place[s] the burden on the disobedient party to avoid expenses by showing that his failure is justified or that special circumstances make an award of expenses unjust.” *Id.* OpenAI cannot meet that burden here.

Courts also have inherent authority to award attorneys’ fees as a sanction for litigation misconduct, including discovery abuse. *See Chambers*, 501 U.S. at 45-46 (quotations omitted)

(courts possess inherent power to “assess attorney’s fees when a party has acted in bad faith, vexatiously, wantonly, or for oppressive reasons”); *Residential Funding Corp. v. DeGeorge Fin. Corp.*, 306 F.3d 99, 106–07 (2d Cir. 2002) (“Even in the absence of a discovery order, a court may impose sanctions on a party for misconduct in discovery under its inherent power to manage its own affairs.”). OpenAI’s conduct—including several years of misrepresentations about its technical capabilities, the destruction of [REDACTED] of user conversations, and the resulting need for extensive motion practice—amply warrants the exercise of this authority.

News Plaintiffs have been forced to expend hundreds of hours of attorney and expert time that would not have been necessary but for OpenAI’s misconduct. These expenditures include:

First, as discussed above, OpenAI produced its training dataset searches for News Plaintiffs’ content *nine months* after it told this Court that it could not conduct such searches. In the intervening nine months, News Plaintiffs spent an enormous amount of time and money dealing with the training inspection “sandbox,” including many, many cross-country trips by News Plaintiffs’ counsel and their experts, hundreds of emails and calls with OpenAI, and disruptions and software challenges. None of this time and money would have been necessary had OpenAI produced its training dataset searches as News Plaintiffs had originally requested, *See, e.g.*, Ex. 13 at 8 (RFP No. 22 requesting “Documents sufficient to identify each article of Times Content contained in Defendants’ Training Datasets.”); Ex. 7 at 11-12 (Daily News Plaintiffs’ RFP No. 22 requesting the same).

Second, News Plaintiffs were required to re-depose Mr. Monaco because he was inadequately prepared on the topics for which he was designated. Mr. Monaco testified that he only prepared [REDACTED]

[REDACTED] *See* Ex. 24 at 73:8-13 [REDACTED]

[REDACTED]

[REDACTED] *id.* at 42:14-23 [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Under Rule 37(d), an inadequately prepared 30(b)(6) witness is tantamount to a failure to appear, warranting sanctions including attorneys’ fees. Fed. R. Civ. P. 37(d)(1)(A)(i); *Eid v. Koninklijke Luchtvaart Maatschappij N.V.*, 310 F.R.D. 226, 228 (S.D.N.Y. 2015) (quotations omitted) (“Courts treat the production of an unprepared Rule 30(b)(6) witness as tantamount to a failure to appear.”).

Third, News Plaintiffs have had to file multiple motions to compel and letters addressing OpenAI’s failure to preserve and produce output log data and regurgitation analysis datasets. These include the January 13, 2025 letter motion to compel preservation, the March 25, 2025 request for the Court to order disclosure of deletion scope, the June 10, 2025 motion to compel production of regurgitation datasets, and the instant motion—none of which would have been necessary had OpenAI complied with its discovery obligations.

Fourth, News Plaintiffs invested months of negotiations, briefs, oral arguments, and confidential settlement discussions to secure the 20 million log sampling process—a process that, as Mr. Monaco’s deposition revealed, may not have been necessary because OpenAI had the ability to search its output logs all along. These expenditures of hundreds, if not thousands, of hours of attorney and expert time were the direct and foreseeable result of OpenAI’s misrepresentations about its technical capabilities.

OpenAI cannot show that its conduct was “substantially justified” or that “other circumstances make an award of expenses unjust.” Fed. R. Civ. P. 37(b)(2)(C). “Conduct is substantially justified if there was a genuine dispute or if reasonable people could differ as to the appropriateness of the contested action.” *Klein v. Torrey Point Grp., LLC*, 979 F. Supp. 2d 417, 442 (S.D.N.Y. 2013) (quotations omitted). There is no genuine dispute here: OpenAI misrepresented its technical capabilities, concealed the existence of relevant datasets and search tools, and destroyed [REDACTED] of user conversations while this litigation was pending. Accordingly, an award of fees and costs is mandatory. *See Novak*, 536 F.3d at 178 (quotations omitted) (Rule 37(d)(3) “requires the award of expenses unless the disobedient party meets its burden.”).

E. The Court May Also Impose Each of the Requested Remedies Pursuant to its Inherent Authority.

The fact that OpenAI has violated so many different provisions of Rule 37, as discussed above, should be sufficient in itself to bring the Court’s inherent authority into play. Nothing in Rule 37, however, addresses a situation in which a party misleads a court to the extent that OpenAI has in this case. To recap simply: [REDACTED]

[REDACTED]. News Plaintiffs sought that material in multiple discovery requests. For the next two years, OpenAI withheld that information, denied it was capable of conducting searches over its databases, and then led the Court and the Plaintiffs down a garden path leading to the “sandbox” and the 20 million-log sample of outputs. This “plan of obstruction, delay, harassment, and expense sufficient to reduce [News Plaintiffs] to a condition of exhausted compliance” is improper and sanctionable. *Chambers*, 501 U.S. at 41 (quoting district court). Each of the remedies requested is an appropriate exercise of this Court’s inherent authority and necessary both in response to OpenAI’s misconduct and as a deterrent to those who might

contemplate following their example. *Nat'l Hockey League v. Metro. Hockey Club, Inc.*, 427 U.S. 639, 643 (1976) (describing sanctions as “not merely to penalize those whose conduct may be deemed to warrant such a sanction, but to deter those who might be tempted to such conduct in the absence of such a deterrent.”).

IV. CONCLUSION

For the foregoing reasons, the Court should grant News Plaintiffs’ motion for sanctions.

Dated: July 9, 2026

/s/ Steven Lieberman

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Robert Parker (*admitted pro hac vice*)
Jenny L. Colgate (*admitted pro hac vice*)
Alexandra Hughes (*admitted pro hac vice*)
Mark Rawls (*admitted pro hac vice*)
Michael H. Jones (*admitted pro hac vice*)
Kristen J. Logan (*admitted pro hac vice*)
Bryan B. Thompson (6004147)
Mary L. Mullins (*pending pro hac vice*)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
901 New York Avenue, N.W., Suite 900 East
Washington, DC 20001
Telephone: (202) 783-6040
Facsimile: (202) 783-6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com
rparker@rothwellfigg.com
jcolgate@rothwellfigg.com
ahughes@rothwellfigg.com
mrawls@rothwellfigg.com
mjones@rothwellfigg.com
klogan@rothwellfigg.com
bthompson@rothwellfigg.com
mmullins@rothwellfigg.com

Attorneys for Daily News Plaintiffs

/s/ Ian Crosby

Ian Crosby (*pro hac vice*)
Katherine M. Peaslee (*pro hac vice*)
Susman Godfrey L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 516-3880
Facsimile: (206) 516-3883
icrosby@susmangodfrey.com
kpeaslee@susmangodfrey.com

Davida Brook (*pro hac vice*)
Emily K. Cronin (*pro hac vice*)
Adnan Muttalib (*pro hac vice*)
Susman Godfrey L.L.P.
1900 Ave of the Stars, Suite 1400
Los Angeles, CA 90067
Telephone: (310) 789-3100
Facsimile: (310) 789-3150
dbrook@susmangodfrey.com
ecronin@susmangodfrey.com
amuttalib@susmangodfrey.com

Elisha Barron (5036850)
Zachary B. Savage (ZS2668)
Alexander Frawley (5564539)
Eudokia Spanos (5021381)
Susman Godfrey L.L.P.
One Manhattan West
New York, NY 10001
Telephone: (212) 336-8330
Facsimile: (212) 336-8340
ebarron@susmangodfrey.com
zsavage@susmangodfrey.com
afrawley@susmangodfrey.com
espanos@susmangodfrey.com

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Kristen J. Logan (*pro hac vice*)
Rothwell, Figg, Ernst & Manbeck, P.C.
901 New York Avenue, N.W., Suite 900 East
Washington, DC 20001
Telephone: (202) 783-6040
Facsimile: (202) 783 6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com

klogan@rothwellfigg.com

*Attorneys for Plaintiff
The New York Times Company*

/s/ Matthew Topic

Jon Loevy (*pro hac vice*)
Michael Kanovitz (*pro hac vice*)
Matthew Topic (*pro hac vice*)
Thomas Kayes (*pro hac vice*)
Steven Art (*pro hac vice*)
Kyle Wallenberg (*pro hac vice*)
Shelley Geiszler (*pro hac vice*)
LOEVY & LOEVY
311 North Aberdeen, 3rd Floor
Chicago, IL 60607
312-243-5900 (p)
312-243-5902 (f)
jon@loevy.com
mike@loevy.com
matt@loevy.com
kayes@loevy.com
steve@loevy.com
wallenberg@loevy.com
geiszler@loevy.com

Stephen Stich Match (No. 5567854)
Lauren Sonnenberg (*pro hac vice*)
LOEVY & LOEVY
222 Sutter Street, Suite 600A
San Francisco, CA 94108
(312) 243-5900
match@loevy.com
sonnenberg@loevy.com

*Attorneys for Plaintiffs The Center for
Investigative Reporting, Inc. and The
Intercept Media, Inc.*

/s/ Lacy H. Koonce, III

Lacy H. ("Lance") Koonce, III
Matthew A. Leish
Brendan Kehoe
Lucy Hoffman

Maya Katalan
Klaris Law PLLC
161 Water Street
New York, NY 10038
Telephone: (646) 779-4882
lance.koonce@klaris.com
matthew.leish@klarislaw.com
brendan.kehoe@klarislaw.com
lucy.hoffman@klarislaw.com
maya.katalan@klarislaw.com

Guy Ruttenberg (*pro hac vice*)
Ruttenberg IP Law, A Professional
Corporation
1801 Century Park East
Suite 1920
Los Angeles, CA 90067
Telephone: (310) 627-2270
ZD-AICopyright@ruttenbergiplaw.com
*Attorneys for Plaintiffs Ziff Davis, Inc.,
Ziff Davis, LLC, IGN Entertainment, Inc.,
Everyday Health Media, LLC, Mashable,
Inc., and CNET Media, Inc.*