

P8CJOPE1

1 UNITED STATES DISTRICT COURT
2 SOUTHERN DISTRICT OF NEW YORK

-----x

3 IN RE: OPENAI, INC. COPYRIGHT
4 INFRINGEMENT LITIGATION.

25 MD 3143

5 Conference

6 -----x

New York, N.Y.
August 12, 2025
10:00 a.m.

8 Before:

9 HON. ONA T. WANG,
10 Magistrate Judge

11 APPEARANCES

12 SUSMAN GODFREY LLP
13 Attorneys for Class Plaintiffs and The New York Times
Company

14 BY: DAVIDA BROOK
15 ROHIT NATH
16 ADNAN MUTTALIB
17 ALEJANDRA SALINAS
JORDAN CONNORS
CRAIG SMYSER
CHARLOTTE LEPIC
HENRY WALTER

18 THE NEW YORK TIMES
19 Attorneys for Plaintiffs The New York Times Company

20 BY: DEMETRI BLAISDELL
KAREN CHESLEY

21 ROTHWELL FIGG ERNST & MANBECK, P.C.
Attorneys for News Plaintiffs

22 BY: KRISTEN LOGAN
JENNIFER MAISEL

23 LOEVY & LOEVY
24 Attorneys for Consolidated Plaintiff The Center for
Investigative Reporting, Inc.

25 BY: MATTHEW TOPIC

P8CJOPE1

Appearances Continued

KLARIS LAW, PLLC

Attorneys for Plaintiff Ziff Davis

BY: LANCE KOONCE

-and-

RUTTENBERG IP LAW, PC

BY: GUY RUTTENBERG

CHARLES NAGGAR

LIEFF CABRASER HEIMANN & BERNSTEIN, LLP

Attorneys for Class Action Plaintiffs

BY: RACHEL GEMAN

WESLEY DOZIER

DANNA ELMASRY

-and-

JOSEPH SAVERI LAW FIRM, LLP

BY: CHRISTOPHER YOUNG

-and-

BOIES SCHILLER FLEXNER LLP

BY: JOSHUA STEIN

MORRISON & FOERSTER LLP

Attorneys for Defendant OpenAI Inc.

BY: JOSEPH C. GRATZ

CAROLYN HOMER

ROSE S. LEE

ZACHARY NEWMAN

ERIC NIKOLAIDES

MARIA FERNANDA RODRIGUEZ

-and-

KEKER, VAN VEST & PETERS LLP

BY: MICHELLE YBARRA

JAMES R. SLAUGHTER

THOMAS GORMAN

SARAH SALOMON

ELEANOR BROCK

-and-

LATHAM & WATKINS LLP

BY: ELANA NIGHTENGALE DAWSON

HERMAN YUE

MARGARET GRAHAM

LUKE BUDIARDJO

-and-

BY: KATYA STYNES

SUSAN KIM

MAILE YEATS-ROWE

P8CJOPE1

Appearances Continued

ORRICK, HERRINGTON & SUTCLIFFE
Attorneys for Defendant Microsoft Corporation
BY: ANNETTE L. HURST
LARA NAJEMY
DANIEL JUSTICE
ANDREW SCHREIBER
-and-
FAEGRE DRINKER BIDDLE & REATH LLP
BY: CARRIE A. BEYER
-and-
NICHOLAS MANHEIM

P8CJOPE1

(Case called; appearances noted)

THE COURT: All right. I think this is everybody on my seating chart. Good morning. We're here for another status conference. I think I was planning to go through the joint charts in the order in which they were filed and the order in which they were raised. However, I wanted to say a brief word or two about the spoliation and data sampling issues because they seem to take up a lot of space. And so we'll do that, and then we won't really deal with those issues in the chart today. We'll probably defer discussion of most of those issues on that section of the chart until Friday's confidential session since, as I read the chart, many of the issues relate to questions that I might have about experimental design and what information or what analysis seems to make the most sense for the retained data and to run on the chat logs.

However, I did want to let the parties know that, in addition to the two intervention motions that were filed and then denied at ECF Nos. 76 and 237, the Court has also been receiving correspondence from concerned users of ChatGPT who have largely echoed the points of the proposed intervenors.

Many of the writings I have received continue to write about concerns about permanent retention and the retention or possession by the government of sensitive data without users' consent. This is paraphrasing some of the things that I've seen in the writings. And in particular, they have concerns

P8CJOPE1

1 about surveillance and control by the government of its
2 citizens.

3 Specifically many users are concerned about how, if
4 vast amounts of data – again, paraphrasing and quoting a little
5 bit – how if vast amounts of data are collected and preserved
6 without democratic safeguards, they can be used a by an
7 authoritarian government as an instrument of surveillance and
8 control.

9 And one issue that we haven't discussed in detail,
10 although I believe it was brought up in one of the conferences
11 when we first learned or when it was first brought to her
12 attention that the press plaintiffs were concerned about the
13 automatic destruction of certain chat logs, is whether there
14 are any U.S. or foreign laws and regulations, and in particular
15 the GDPR, that might create tension with the principles, rules,
16 and case law that have developed around document destruction
17 and suspension of those policies in the context of civil
18 litigation.

19 So one thing I am considering – and this is to give
20 you a heads up, but there won't be an order on this issue until
21 after this conversation – one thing I am considering is asking
22 the parties to brief these issues more fully. I am considering
23 also putting some of the correspondence, the concerned citizen
24 correspondence, into a document on the docket that is
25 attorneys' eyes -- visible to attorneys only, not to the

P8CJOPE1

1 general public, so that you all can see some of the concerns
2 that are being raised with me. I don't know if your clients
3 are also receiving similar correspondences.

4 So therefore, one thing I am considering is asking the
5 parties to brief these issues more fully. I'll probably wait
6 until you all have had an opportunity to review some of the
7 correspondences and see how we wish to proceed. I will say
8 generally that OpenAI's customers seem to be under the
9 misimpression that the Court or I personally will be in
10 position of this data or I'm keeping some of it – that could
11 not be farther from the truth – and that this data will be
12 retained permanently or be made public, even though I
13 repeatedly stated that none of these facts is true, and I
14 stated that on the record and I'll say that again.

15 But because it does relate in part to some of the
16 discussions we're having about the data sampling and about data
17 retention generally, I think what I will do is I will put the
18 correspondences on the docket visible only to the parties and
19 to the Court. And then maybe come Friday we can have a
20 conversation about what, if anything, we should be doing about
21 them and whether there should be any further briefing on the
22 issues and particularly the tension with what users believe are
23 their privacy rights and whether their privacy rights are being
24 infringed by the retention order. Okay?

25 All right. Let's move on to Exhibit A. So I'm

P8CJOPE1

1 looking at ECF 437-1, document discovery and inspection issues.
2 So this appears to be OpenAI raising the dispute. It's related
3 to ECF 377 and 376. And I guess my first question for OpenAI
4 is what is it about these lost or destroyed chat logs that is
5 discoverable, and how does it relate to any of the claims or
6 defenses in the case?

7 MS. YBARRA: Thank you, your Honor. If you'll allow
8 me, I'm happy to address your questions and prepared to address
9 those. I want to do a little bit of table setting about how we
10 got here and what exactly these logs contained.

11 These logs are from the Times's use of Chat Explorer.
12 This is an internal tool that the Times used to interface with
13 OpenAI's models. That is distinct from the Times's broader use
14 of generative AI applications or tools, but really the Times's
15 use of defendant's own products.

16 That discovery is directly relevant to our defenses
17 under the fair use factors, our contributory infringement
18 defense, and it's relevant to rebutting the Times's arguments
19 that OpenAI's models are copyright regurgitation machines that
20 offer nothing innovative or new. The Times's own use of our
21 products squarely rebuts those arguments.

22 THE COURT: Is it the fact of the use or is it somehow
23 something they're doing in the logs? Because I don't think
24 it's disputed that the Times used Chat Explorer and used
25 ChatGPT, right?

P8CJOPE1

1 MS. YBARRA: Correct, your Honor. It is the content
2 of what is happening in the logs.

3 THE COURT: Okay.

4 MS. YBARRA: The records of the Times's own use are
5 directly relevant to the first fair use factor because they'll
6 show that OpenAI's technology is transformative, that OpenAI's
7 platform provides new and innovative tools that serve a
8 different purpose than the Times's asserted works. This is
9 exactly the kind of evidence the Supreme Court relied on in
10 *Google v. Oracle* in finding fair use.

11 And specifically, we don't have a lot of evidence
12 about the content of those logs because the evidence was either
13 deleted or the Times refused to produce the remaining log
14 evidence. But there's one document, it's attached as
15 Exhibit 18 to our motion. And this appears to be a Times
16 employee who is independently tracking his prompts and outputs
17 to Chat Explorer. And you can see he was using OpenAI's
18 models, for example, to do sentiment analysis, meaning he was
19 using it to analyze a set of data to determine the emotional
20 tone and opinion of a large set of data.

21 And businesses do that when they might want to
22 understand public reaction to a particular product or
23 initiative. And for an online publisher like the Times where
24 audience trust and engagement are critical, this is going to be
25 very valuable, allows them to analyze and respond to audience

P8CJOPE1

1 feedback at scale rather than just paying attention to the
2 loudest voice in the room like a couple of reviews or comments.
3 So sentiment analysis is a great example that's entirely
4 distinct from the purpose of the original works, which was to
5 inform and entertain.

6 Another example from that same exhibit, which is just
7 a handful of prompts and outputs, this Times employee was
8 assessing readability and scoring readability to note the
9 content was properly tailored to the audience the Times was
10 attempting to reach. We also expect that the other logs that
11 the Times is withholding would show similar conduct and
12 additional transformative uses like analyzing large data sets,
13 doing data crunching to assist in investigative efforts, those
14 sorts of things. All of those are directly probative of
15 transformative use and relevant under the first fair use
16 factor, including whether OpenAI's models are innovative tools
17 that further the creation and dissemination of creative works,
18 as discussed in *Google v. Oracle*, by helping the Times bring
19 its own works to market.

20 This same data, the same content of those prompts and
21 outputs, also relevant to OpenAI's defense under the fourth
22 fair use factor. The Times's use of our products to further
23 their business efforts to prove their journalism to their
24 advertising, their *Wirecutter* recommendations, however they
25 were using our products to do that, those uses will show that

P8CJOPE1

1 OpenAI's products benefit the Times, and those benefits have to
2 be considered as part of the factor four analysis. Whether the
3 copyright holder benefits from the use and especially where
4 that use aligns with their historical business model, that
5 benefit is relevant to the market effect analysis.

6 So those are just two examples from the fair use
7 inquiry that we think the content of these prompts and outputs
8 are directly relevant to.

9 THE COURT: Okay. I still am having trouble I guess
10 getting to why the content of those logs, which by the way is
11 different from the sampling methodology that we've been talking
12 about for the OpenAI destroyed logs, right, this -- you're
13 getting to the actual prompts and outputs and you want to see
14 the content of all the prompts and outputs. I guess I'm just
15 still trying to understand how if you have an admission and
16 knowledge that this is what the Times is doing, why isn't that
17 enough?

18 MS. YBARRA: Your Honor, we don't have a specific
19 admission about any of this. I referred you to Exhibit 18 that
20 appears to be a one-off happenstance inclusion in the Times's
21 production of the content of the logs. But the Times was using
22 this product starting in October of 2023 and onward through --
23 the kind of critical period was through March 2024 when this
24 Chat Explorer product was shelved in favor of a different GenAI
25 product, and we don't have any of those documents. Those are

P8CJOPE1

1 the documents that were not preserved. And the Times has some
2 remaining logs from after that product was shelved, but is
3 refusing to produce those.

4 To your question -- so we don't have an admission from
5 anybody yet on what the Times was doing. We're not able to
6 question witnesses about their use without those remaining logs
7 and --

8 THE COURT: I mean, why can't you just ask them the
9 question?

10 MS. YBARRA: Your Honor, those questions in that
11 inquiry are going to be much more powerful with the records to
12 refresh recollections and show the witnesses what was
13 happening. I expect we're going to get a lot of "I don't
14 knows" or "I don't remembers" in response to just open-ended
15 questions without evidence.

16 In addition to that, you asked about why is it the
17 fact -- why do you need content? Summaries are not the same.
18 The fact that we know they used the logs or the Times has
19 pointed to, well, we've produced the guidelines for Chat
20 Explorer, why isn't that enough? That's not the most probative
21 evidence of transformative use. Summaries are not the best
22 evidence. They flatten the nuance of the use, and it will
23 allow the Times to abstract away the true transformative nature
24 of OpenAI's models and the benefits the Times gleaned from
25 them.

P8CJOPE1

1 So one way to think about it would be like a
2 use-of-force police report recording a use-of-force incident.
3 You get a short summary of what happened versus the actual body
4 cam footage showing you what happened in the moment or reading
5 the inside of a book jacket to understand what the book was
6 about versus actually reading the book. The Times's actual
7 prompts and OpenAI's outputs are going to be the most probative
8 evidence on transformative use in particular.

9 In addition, the Times has repeatedly committed to
10 producing evidence to OpenAI about its use of OpenAI's
11 products. It's done that for months throughout this
12 litigation. Your Honor has relied on it. Judge Stein has
13 relied on those representations by the Times and has done so in
14 order to justify denying OpenAI other discovery. This is the
15 core evidence of the Times's use of OpenAI's models, and we
16 think it should be produced.

17 THE COURT: Okay. Let's hear from the Times.

18 MS. BROOK: Good morning, your Honor.

19 Your Honor began with the question, defendants began
20 with what they called table setting. I will go straight to the
21 question. But just for the record, I want to note we disagree
22 with much of what was said factually in defendants' remarks.
23 But to begin with the question, what is it about these logs
24 that is discoverable and how does it relate to any of the
25 claims in this case, it does not. It is exactly as your Honor

P8CJOPE1

1 put it. It is the fact of the use. It's not the specific
2 inputs, and the specific outputs in this context that
3 defendants need in order to make their case.

4 In three briefs, defendants – and this is an issue
5 also raised by Microsoft, your Honor, the next one on the chart
6 – in their briefs, defendants focus on three reasons why this
7 might be relevant. In their argument they narrowed it down to
8 one, which makes sense, because I think the other two have
9 already been ruled on by this Court, as well as Judge Stein.

10 But just for the record, the first one that they claim
11 is noninfringing use, which Judge Stein has made clear is not a
12 defense at issue in this case. The last one is manipulated
13 regurgitation, the good old Exhibit J example, which this Court
14 has dealt with more times than I care to recall. But as for
15 the fair use one, it seems like we're zeroing in on that, and
16 we're zeroing in on two particular reasons why it might be
17 relevant fair use.

18 The first one is that OpenAI argues that this evidence
19 would support its finding under factor four that this somehow
20 shows that the Times is benefiting from OpenAI's product. I
21 would contend that that distorts the factor four fair use
22 analysis beyond recognition a little bit. The factor there is,
23 "Whether the copy brings to the marketplace a competing
24 substitute for the original or its derivative so as deprive the
25 rights holder of significant revenues because of the likelihood

P8CJOPE1

1 that potential purchasers may opt to acquire the copy in
2 preference to the original." That's from *Authors Guild v.*
3 *Google*.

4 That's what we're looking at there under the fair use
5 analysis. And of course the Times's own use even of ChatGPT
6 itself is not relevant to that analysis precisely because the
7 *New York Times* pays to use ChatGPT, which is the critical thing
8 that it is of course asking OpenAI and Microsoft to do when it
9 comes to building its models. It's not saying don't do it;
10 it's saying please pay for it. So that that is the first one.

11 The second one, it argues that this is maybe somehow
12 relevant to fair use because I guess it shows the
13 transformative nature. And I guess again, your Honor, the
14 Times has admitted that it uses ChatGPT. It has produced ample
15 discovery on this, and the representations that it hasn't or
16 that it can't get at this information via any mechanism other
17 than these logs is just not correct.

18 OpenAI literally served requests for admissions asking
19 how we use its products. Admit that you use Chat Explorer.
20 Our answer, admitted. OpenAI has served RFAs on the sentiment
21 analysis that Ms. Ybarra was discussing. And again, they
22 served RFAs on that and we answered. They can use emails to
23 show sentiment analysis and readability issues. Again, why do
24 they need the logs? The answer is they don't.

25 It's 80,000-plus logs that would have to be gone

P8CJOPE1

1 through one by one. If the Court is interested in the burden
2 argument, I'm happy to articulate it. But otherwise it was
3 laid out in our briefs that the Court has clearly familiarized
4 itself with, so I will leave it there unless there are any
5 questions.

6 THE COURT: I guess one thing I would ask defendants
7 to explain is how -- one year's worth of logs is 80,000 logs?

8 MS. BROOK: It's 16 months' worth of logs, your Honor.

9 THE COURT: What's the prejudice? Because as I
10 understand it, there's logs that are retained before and after.
11 It's just this 16 months when the logs were transferred to this
12 other repository.

13 MS. BROOK: It's quite the opposite. So the way that
14 they describe it, they do make it seem as though the bulk of it
15 wasn't retained. It's exactly the reverse. So it's roughly
16 three months' worth of logs that weren't properly preserved
17 despite best efforts. And we have articulated that to them and
18 explained it and put it in a sworn affidavit on the record
19 explaining what happened there.

20 But everything has been preserved since I believe
21 March 24, 2024, which is about 16 months and counting, worth of
22 logs. And that's what they're asking for now, both OpenAI and
23 Microsoft, is those 16 months' worth of logs. Again, they say
24 that it's been -- I don't recall the exact word they used that
25 that's been stopped to be used or turned off or something like

P8CJOPE1

1 that. That's not accurate either. Indeed, I think I checked
2 before this hearing, and just in the last two months there are
3 an additional 20,000 log entries that we've captured and
4 preserved.

5 So the use is ongoing. It's going to go on forever.
6 And there's again just no reason why it would make any sense in
7 terms of the issues in this case for The New York Times to have
8 to go through those 80,000 logs and produce them one by one.

9 THE COURT: Okay. Let me ask Ms. Ybarra or anybody
10 else from OpenAI who wants to speak on this issue. Under what
11 rule, what federal civil rule, are you seeking to get these
12 logs? Because when I read it, I read it with the overlay of a
13 spoliation argument. That's not what it sounds like you're
14 asking for.

15 MS. YBARRA: No, your Honor. These are RFPs. These
16 logs are responsive to multiple RFPs that we served, and I've
17 got about ten of them that I can identify for you. So this is
18 not something outside of the normal discovery process under the
19 FRCP.

20 I'd like to respond to your questions and a few points
21 that Ms. Brook raised. First, she said that the argument that
22 these are relevant to OpenAI's defense on contributory
23 infringement because they show noninfringing uses. Counsel
24 argued that that has been decided by Judge Stein already and
25 that argument is foreclosed to us.

P8CJOPE1

1 I don't think that's accurate. I believe counsel was
2 referring to Judge Stein's order on the motion to dismiss.
3 That order on the motion to dismiss does not foreclose
4 discovery here. The order merely said that at the motion to
5 dismiss stage, the existence of substantial noninfringing uses
6 do not foreclose plaintiff's contributory infringement claims,
7 saying nothing about their discoverability.

8 Second, Ms. Brook talked about how the transformative
9 nature isn't really at issue or these aren't relevant to
10 transformative use because the Times admits that it uses
11 ChatGPT. Sure. But the Times does not admit that ChatGPT or
12 OpenAI's models are transformative. In fact, your Honor was at
13 the technology tutorial in June, and I'm sure heard the news
14 plaintiffs stand up and say OpenAI's technology is nothing
15 innovative, nothing new. It's a regurgitation machine that
16 spits out other people's content.

17 Third, the burden issue I would like to address,
18 because counsel just told the Court and it's in the Times's
19 papers that we're talking about 80,000 logs. Correct. The
20 80,000 logs can be exported into an Excel document. The
21 average length of a user prompt to ChatGPT is 350 characters.
22 That's a couple of sentences. What we're talking about in
23 terms of the burden of review and production here is minimal.
24 That's less than two gigabytes of data. And this is not
25 something that's going to require an extensive privilege

P8CJOPE1

1 review.

2 The Times's own answers to our preservation questions
3 about this data – this is Exhibit 1 to our motion – shows that
4 no lawyers, the legal department, did not use Chat Explorer.
5 The Times's policies around the use of that internal tool did
6 not allow employees to put in sensitive information that could
7 identify people. This is not a burdensome review. This is
8 something that could be conducted quite quickly.

9 And finally, to the Court's question about it seems
10 like 16 months of data were preserved. That is correct, but
11 the Times has refused to produce it. To be clear, there was a
12 few months of data that we believe was the core period of the
13 Times's experimentation with OpenAI's models. And we've cited
14 documents. I'm not going to read them into the record here
15 because the Times has designated them confidential. But their
16 use of OpenAI's tools dropped off dramatically in March 2024
17 and onward.

18 In addition, we understand that six custodians, six
19 document custodians who are going to be deponents in this case
20 and witnesses in this case, were using this tool in the
21 critical experimentation period. Their logs have been
22 destroyed. Only two of those six custodians have any records
23 from the remaining 16 months of data, and we still don't have
24 those. That's why we're asking for production of the entirety
25 of the remaining Chat Explorer data, which we don't believe

P8CJOPE1

1 will be burdensome at all.

2 And I'll stop there unless the Court has any
3 questions. I know my colleague from Microsoft is eager to
4 address this issue as well.

5 THE COURT: Okay. Well, let's talk about the
6 proportionality factors. Why don't we let Microsoft have a
7 word, and then I'll turn it back to the New York Times. I know
8 you didn't get to the burden. It's really not just burden, but
9 we have to talk about proportionality as a whole. And I
10 encourage you not to go over anything that you've already put
11 in your briefing, and let's hear what you have to say.

12 MS. YBARRA: Your Honor, I'm happy to walk through
13 those factors if you'd like to right now, or if you'd like to
14 hear from Microsoft we can do that too.

15 THE COURT: Let's hear from Microsoft because they
16 haven't spoken yet, and we'll just hear from everybody for a
17 bit.

18 MR. JUSTICE: Before going to the proportionality, I
19 think it's important when addressing proportionality to put it
20 in the context of relevance. I don't want to repeat arguments
21 that Ms. Ybarra has already made, but I want to emphasize the
22 defense of substantial noninfringing uses.

23 So Count Four of the second amended complaint is an
24 end user contributory copy infringement claim. And substantial
25 noninfringing uses is a defense to such a claim, especially if

P8CJOPE1

1 the claim is based on material contribution, which is the case
2 here. It's also relevant that the New York Times is an end
3 user of the product.

4 So the actual content of these logs and whether or not
5 they show infringing uses or noninfringing uses is evidence of
6 whether or not there's substantial noninfringing uses by the
7 products.

8 THE COURT: Wait a minute. How -- I'm not sure -- I'm
9 trying to understand where you're going with this and
10 whether -- and clearly what the defense is, like what is the
11 claim? And what is the defense? And what must be shown? Is
12 it that the plaintiff -- are you claiming that The New York
13 Times is infringing its own copyrights or is able to use
14 ChatGPT without infringing its own copyrights, and therefore
15 that's a defense? I'm not really clear what you're saying
16 here.

17 MR. JUSTICE: The content of the prompts and outputs
18 would show that there's no infringing content or no infringing
19 outputs by the product.

20 THE COURT: Okay. Let's pause right here because I
21 seem to recall a conference in the very, very early days of
22 this case where someone on the plaintiffs' side, it may have
23 been Ms. Maisel, but I don't want to misattribute, when we were
24 talking about output can also be infringing, right? And we're
25 talking about the plaintiffs' claims that there's potentially

P8CJOPE1

1 three stages at which infringement is alleged to have happened.

2 And so I guess I'm first of all trying to understand
3 which stage or if it's all three stages that you're saying that
4 the substantial noninfringing use is a defense. And then drill
5 down a little bit more. Whose substantial noninfringing use,
6 and what substantial noninfringing use matters?

7 MR. JUSTICE: So the substantial noninfringing use is
8 a defense in particular to the New York Times's Count Four,
9 which is the end user contributory copyright infringement
10 claim. There are two general flavors of that type of claim.
11 One is conduct that encourages the infringement, so it's the
12 inducement type. And the second is the provision of a
13 technology that facilitates the infringement. That is the
14 material contribution type claim. The New York Times is
15 asserting the latter, the material contribution claim. And
16 under *Sony* and *Grokster* and *Matthew Bender* --

17 THE COURT: When you say The New York Times is
18 asserting the latter, a material contribution type of claim,
19 you're saying that their claim is that the ChatGPT does what?

20 MR. JUSTICE: That the outputs are infringing and that
21 because the defendants have created a technology that allows
22 the end user to engage in infringing activity, the defendants
23 are contributorily liable for that end user's infringement.

24 THE COURT: Okay. So why does the fact that anybody
25 including the plaintiff can use this technology to generate

P8CJOPE1

1 output that is infringing, why -- I mean, it's -- I seem to
2 recall -- this is going back to what I was saying before I
3 asked you to drill down into the substantial noninfringing uses
4 defense. I seem to recall that in one of our early, early
5 conferences somebody on plaintiffs' side, when they were
6 talking about outputs being infringing, said substantially that
7 the results of some of these conversations generate infringing
8 output, right, that either word for word -- regurgitation we'll
9 call it, regurgitation of copyrighted material.

10 And I seem to remember somebody from the defendants'
11 side saying, well, what we'll find is that only you'd have to
12 work really, really hard, any end user would have to work
13 really, really hard to get that infringing output out. And
14 there was at least a suggestion or maybe just an inference that
15 I drew that what you proposed to show or thought or sought to
16 show was that New York Times users were the only ones who were
17 trying to get around the protections against generating
18 copyrighted material and to produce the copyrighted material.

19 Now I'm hearing that you want these logs because you
20 want to show that there's substantial noninfringing uses. So
21 this is partly why I'm a little bit confused right here.

22 MR. JUSTICE: I think your confusion may be coming
23 from the fact that up until now, the direct infringement claims
24 have been discussed more in previous hearings. But there is
25 this Count Four, which is pled in the alternative, which says

P8CJOPE1

1 that if end users are using it for copyright infringement
2 purposes then the defendant should be held contractually
3 liable, and it is the contents of the logs that will show the
4 various pro-copyright noninfringing uses that the models are
5 capable of.

6 And in the type of claim they brought here, the
7 contribution claim, not the inducement claim, as explained in
8 *Sony* and *Grokster*, what the Court looks to is as long as
9 there's no inducement act where the defendants are attempting
10 to get the end user or motivating them to commit infringement
11 but they are just providing a technology that some users could
12 use for infringement while other users don't, then it's
13 essential to get the evidence of all the different types of
14 noninfringing uses and the extent of the noninfringing uses.

15 So here, the Times in their RFAs, which is Exhibit I,
16 they admit just to they use the GPT models for research, for
17 brain storming and editing articles. But without actually
18 looking at the logs themselves, we can't say what other uses
19 they're using and exactly how they're using it.

20 The documents they produced, the guidelines about how
21 to use it, some documents that are summarizing how Times
22 employees are using it are circumstantial, and the only direct
23 evidence of the use is the logs themselves is by far the best
24 evidence of how it is actually being used. It's like looking
25 at an actual deposition of record versus some summary of a

P8CJOPE1

1 deposition.

2 As to burden, so that's why it's relevant under
3 substantial noninfringing uses defense to Count Four. Under
4 the burden, this information is not custodial information.
5 They're not going to have to collect it from the 58 different
6 individuals. It's in one central database. It's less than two
7 gigabytes of data. Microsoft has during meet and confer made
8 an offer to the Times that it can remove all of the logs for
9 its in-house attorneys, which would greatly reduce the chance
10 of any likelihood that there be any privileged information in
11 these logs.

12 We don't see a likely case where a non-attorney would
13 be putting in attorney-client communications into GPT. This
14 is less than two gigabytes of data they could export onto a
15 drive and produce to us. It's about 82,000 entries. You only
16 have to look at the prompt itself to do a privilege review or
17 to do a kind of generalistic privilege review, which would
18 be -- a prompt is generally a sentence, a couple of sentences,
19 so we're not talking about a very burdensome review for
20 privilege or any other reason.

21 (Continued on next page)

P8CEOPE2

1 THE COURT: All right. Let's hear from someone else,
2 whether Microsoft wants to talk about --

3 MS. BROOK: That was Microsoft.

4 THE COURT: Oh, that was Microsoft. Okay.

5 Go ahead, Ms. Brook.

6 MS. BROOK: You've got both OpenAI and Microsoft, so
7 let me see if I can handle both.

8 So, beginning --

9 THE COURT: All right. There's a lot of noise there,
10 there's cords on the floor. Please be careful not to disturb
11 them.

12 THE DEPUTY CLERK: It's actually you, sir. It's very
13 poor planning on where the cords are, so apologies.

14 THE COURT: Unfortunately. Sorry.

15 THE DEPUTY CLERK: If you hit it, it makes noise.

16 MS. BROOK: So, your Honor, let me briefly address the
17 principle arguments by OpenAI and Microsoft and then I'll turn
18 to the factor for proportionality, if it pleases the Court.

19 THE COURT: Yes.

20 MS. BROOK: So, on this non-infringing use debate,
21 again, I posit the confusion here, which is what Judge Stein
22 called out in his order on the motion to dismiss in this case,
23 is that, yes, we have a claim involving contributory
24 infringement, but it is not contributory infringement by
25 inducement. We do not have a claim by inducement.

P8CEOPE2

1 And to the extent the evidence that counsel for
2 Microsoft was contending is relevant at all, which I don't
3 think it -- it's certainly not the best evidence, and I don't
4 think it's relevant really at all, it would go to that
5 inducement prong as he articulated, not to just a good
6 old-fashioned contributory infringement claim, which is what is
7 at issue in this case.

8 Going back to just what it seems like they want to get
9 at, which is how The Times uses these technologies, again,
10 we've produced considerable evidence, data on all of these
11 facts, and they haven't taken a single deposition. Let them
12 start asking our people questions, "Well, how, if at all, do
13 you use it?" And get the answers. And that is, I would say,
14 the best way, let alone the most proportional way, for them to
15 get at the additional color that they seek here.

16 But turning to proportionality, your Honor, this is
17 where I will do some table-setting, because I do think it's
18 very important and it's what the courts contemplated when they
19 adopted the new federal rules here. We have to look at this in
20 context. I'm not saying this critically, your Honor, I am
21 saying this factually. Microsoft served, to date, 282 requests
22 for production on The New York Times alone. OpenAI served 316
23 requests for production on The New York Times alone. Microsoft
24 followed up with 67 requests for admission. OpenAI has 125
25 requests for admission. Which is just to say, your Honor, that

P8CEOPE2

1 there is a lot of discovery that they have requested in this
2 case.

3 And okay, fine. Again, I'm here factually, not
4 critically, though I reserve the right to be critical at
5 another point in time. And within that context, your Honor,
6 this is just not what the parties should be focused on. We
7 have made incredible efforts to produce troves of data,
8 including on these very issues of The Times' use of the ChatGPT
9 products. We did that without any court intervention or order
10 from the Court, it was all voluntary, and they have that data,
11 and they can follow up with it as they see fit now as we enter
12 the depositions phase of this case.

13 But the additional information that they're requesting
14 here, 80,000 files -- so, here I might make an error, your
15 Honor, so apologies, but I'm thinking back to the Court's
16 back-and-forth with my colleague, Mr. Frawley, where we were
17 going back and forth over how many documents OpenAI and
18 Microsoft should have to even review, your Honor; right? The
19 search terms. How many documents should they even have to look
20 at related to the whole custodial database. And I think the
21 number we arrived at was 600,000. So, they want to add an
22 additional 80,000 documents to our plate at this stage of the
23 case?

24 And whether they're short or long, I don't know. I
25 think they would agree it can vary, but we will have to review

P8CEOPE2

1 every single one of those prompts one by one in an additional
2 80,000 documents to see whether or not they have information
3 that should not be produced either because it is privileged --
4 and sure, it's not just lawyers that talk sometimes about
5 privileged information. You can look to the privilege logs
6 defendants will serve in this case for that fact. And it's
7 just not privilege here, your Honor, it's the reporter's
8 privilege. Right? We have to look at whether or not every
9 single one of these logs maybe involved some conversation that
10 is protected by that, which is going to require, I fear, a very
11 significant amount of back and forth with individual reporters
12 who, as this Court has appreciated in the past, are rightfully
13 very sensitive to these types of inquiries and this type of
14 information gathering. So, at bottom, it's just not the best
15 evidence for what they're trying to get at.

16 On top of that, they already have good evidence for
17 what they're trying to get at. And then on top of that, we are
18 about to enter into a new stage of this case where they can get
19 additional evidence in this regard. But I contend that in a
20 case where there are this many requests serve and The Times has
21 already produced as much as it has, and where defendants have
22 pushed back about even reviewing more than 600,000 custodial
23 documents, seeking 80,000 documents that have no tethering to
24 any custodians, no tethering to any search terms, no tethering
25 to any of the time parameters that the parties agreed to in

P8CEOPE2

1 this case, is not proportional to the needs at hand.

2 THE COURT: All right. Anybody else want to add
3 anything?

4 MS. YBARRA: Yes, your Honor.

5 Michelle Ybarra from Keker, Van Nest & Peters for
6 OpenAI.

7 First, to address counsel's comments here, discovery
8 on all sides has been extensive. The Times itself has
9 propounded hundreds of RFPs on OpenAI, have brought
10 interrogatories, RFAs in similar scope to what you just heard.

11 The News plaintiffs, as your Honor is well aware, are
12 also seeking 120 million logs from OpenAI. And I want to
13 correct Ms. Brooks' characterization of what is at issue here.
14 This is not 80,000 documents. As I said before, a prompt is
15 typically a couple of sentences, and a response is usually a
16 paragraph or two. This is data that can be exported into a CSV
17 file. It's not 80,000 documents, it is one centralized
18 database. And The Times that not substantiated any burden with
19 reviewing it.

20 THE COURT: What about the reporter's privilege and
21 the attorney-client privilege topic?

22 MS. YBARRA: Yes, your Honor.

23 So, The Times' own internal documents suggest that
24 those should not be at issue here. First, on the reporter's
25 privilege, as I mentioned before, The Times' internal chat

P8CEOPE2

1 explorer policy prohibits inputting information that could
2 personally identify an individual, such as a confidential
3 source. And you can see that policy. This is attached as an
4 exhibit to Microsoft's motion. I believe it's Exhibit B at
5 page two.

6 On the legal review, you heard Mr. Justice from
7 Microsoft relay the offer made to The Times in meet and confer
8 to remove any in-house lawyers who happen to be using the tool.
9 Now, The Times' own answers to our preservation questions
10 regarding the status at Exhibit 1 to our motion suggest that
11 lawyers were not using the tool. The legal department was not
12 experimenting with the tool.

13 So, we think the burden claims, you'll not, there's
14 nothing in The Times' brief that talks about how many pages
15 this is to review, how many documents it is, how many attorney
16 hours it would take to conduct this review, and we think the
17 burden is not only minimal but it's definitely proportionate to
18 the needs of the case. This evidence, as we talked about
19 before, is directly probative of the key defense in the case,
20 the fair-use defense to copyright infringement.

21 And I, again, refer your Honor to *Oracle v. Google*.
22 There, the Supreme Court said that Google's use of the Sun Java
23 APIs to create new products and expand the usefulness of the
24 platform provided a highly creative and innovative tool that
25 was consistent with which creative process that the basic

P8CEOPE2

1 constitutional objective of copyright itself. The one document
2 we have showing what's in these logs, which is, again, attached
3 at Exhibit 18 to our motion -- this is exactly the kind of
4 evidence that is in The Times' possession -- this is akin to a
5 sort of party admission about the transformative nature of
6 ChatGPT and OpenAI's models.

7 And there's nothing, your Honor, just in terms of
8 trial themes in rebutting the story that the Times intends to
9 tell at trial, as we saw previewed during the technology
10 tutorial, that this ChatGPT is nothing new, it's a
11 regurgitation machine, it doesn't provide an innovative new
12 tool for the world, we are entitled to rebut that, and there's
13 no higher praise than your adversary's adoption of your own
14 technology to further its business interest. That's validation
15 of the fact that we had something new and different to The
16 Times' works, and we're entitled to that evidence.

17 THE COURT: Okay. Anything else?

18 Mr. Justice.

19 MR. JUSTICE: Yeah. Just one quick point.

20 On the policy points and how it would be a violation
21 of The Times' policies to put in confidential sources or
22 publish information, that instruction, those policies, are also
23 shown on the chat explorer interface, so every time someone
24 uses chat explorer they're reminded of those policies. That's
25 in Exhibit 3 to our motion.

P8CEOPE2

1 I would also just like to point out that their entire
2 claim of burden rests on the assumption that their employees
3 are violating this policy, and they have not shown any --

4 THE COURT: No, I don't think so. I mean, that's --
5 like, let's take an example of a regular old privilege log.

6 MR. JUSTICE: Yep.

7 THE COURT: Right? You still have to review -- just
8 because it came from an attorney doesn't mean that it's
9 privileged or not; right? It helps you in the first
10 separation. And just because there's not an attorney on it
11 doesn't mean that there are two non-attorneys talking about
12 advice that they got inside the company. Of course it has to
13 be reviewed. What are you suggesting here?

14 MR. JUSTICE: I'm just saying it's unlikely --

15 THE COURT: You're suggesting they just produce it and
16 not --

17 MR. JUSTICE: No, they have to --

18 THE COURT: -- review it because you assume that the
19 journalists are following their own policies?

20 MR. JUSTICE: I'm just saying that they haven't shown
21 that there is any likelihood of this happening, I'm not saying
22 they shouldn't review --

23 THE COURT: But we're talking about the burden, not
24 that they haven't shown. Do they have to show that there's a
25 likelihood of something happening in order to say that they

P8CEOPE2

1 need to review everything before it goes out?

2 MR. JUSTICE: No, they have the right to review.

3 THE COURT: Right. So that's what we're talking
4 about.

5 MR. JUSTICE: Yeah.

6 THE COURT: Okay?

7 MS. BROOK: The Court made some of my rebuttal for me
8 so I'll be very brief.

9 First of all, on *Oracle*, it gets thrown around a lot.
10 I will refer this Court to docket number 592, which was
11 Judge Stein's order reviewing a series of this Court's
12 discovery rulings that were then objected to, appealed to
13 Judge Stein, where he noted the limits of *Oracle* and that
14 *Oracle* does not stand for the proposition that evidence of the
15 plaintiff's activities is always relevant.

16 On the burden, yes, of course it's true that the
17 policy is probably not followed perfectly and so we need a
18 review for that purpose. But even if the policy is followed
19 perfectly, reporter's privilege isn't just about the source,
20 reporter's privilege is a huge canon of issues that has to be
21 reviewed incredibly carefully, and to the proportionality
22 point, requires an intimate understanding of the individual
23 article that the reporter was writing about at the time in
24 order to understand whether or not this shows The Times'
25 internal news gathering processes. That's not source but that

P8CEOPE2

1 is protected.

2 On the legality, on the legal privilege, the Court hit
3 the nail on the head, which is exactly that we appreciate that
4 they're not going to require that we look at lawyers, but just
5 because there aren't lawyers, that does not mean that folks
6 within The Times at a period when The Times was involved in
7 some of the biggest litigation it's been involved with in
8 decades, were not maybe using this tool for issues related to
9 work product and legal advice, and so we would have to review
10 one by one for that reason as well.

11 THE COURT: Anybody else have anything to add?

12 I see Ms. Ybarra standing up.

13 MS. YBARRA: Michelle Ybarra for OpenAI.

14 Your Honor, I'll be very brief.

15 Counsel referred you to Judge Stein's order and
16 suggested that it discusses the limits of *Oracle* in a way that
17 would make this discovery irrelevant. To be clear,
18 Judge Stein's order on this point denied OpenAI, as did your
19 Honor, discovery into The Times' use of other generative AI
20 products besides OpenAI's. And in reaching that conclusion,
21 the Court relied on and quoted The Times' representation that
22 it had already agreed and already committed to producing
23 discovery about The Times' use of OpenAI's products.
24 Judge Stein was not saying that this isn't relevant. In fact,
25 he specifically cited The Times' commitment to produce the very

P8CEOPE2

1 discovery that is at issue here that we've been waiting for,
2 for months.

3 Thank you, your Honor.

4 THE COURT: All right. I'm going to give you a chance
5 to do some supplemental briefing and then I'm going to write on
6 this one, so I'm not going to rule from the bench.

7 All right. Let's have the supplemental briefing due
8 Friday, close of business. It should cover the following
9 topics:

10 Relevance. And particularly, I want everybody to be
11 very clear on which claim, which defense, and how it relates
12 to, in particular, it sounds like the fourth fair use factor is
13 the one that we're really arguing about. Right? Or is it --

14 MS. YBARRA: Your Honor, Michelle Ybarra for OpenAI.

15 It's factor one and factor four, I believe.

16 THE COURT: Okay. Break out those arguments
17 separately. Cite case law with parentheticals explaining how
18 that case law relates. Okay? Don't just give me the top line,
19 you know, something that I can find elsewhere. Relate that
20 case to the facts here, the circumstances here.

21 Proportionality. In and proportionality, let's go
22 back to Rule 26 and address the particular factors that Rule 26
23 encourages the parties and the Court to consider when making a
24 proportionality decision. I want to hear specifically why the
25 content of the chats matters so much, why it isn't just that

P8CEOPE2

1 The New York Times can answer admissions or answer questions at
2 depositions that go to how The New York Times employees use
3 chat explorer generally.

4 And then in connection with the burden and the expense
5 of the proposed discovery, a little bit more about the scope of
6 the privilege and the scope of the privilege review. I heard a
7 little bit from, Ms. Brook and I will admit I am not as
8 familiar with the reporter's privilege as I am about
9 attorney-client privilege, so I'd like to hear about that and
10 get some understanding about why or how this doesn't -- I don't
11 even want to use the word "infringe" but I'm at a loss for
12 another word right now -- but the tension between the discovery
13 that's sought and potentially the reporter's privilege as well
14 as the attorney-client privilege.

15 This is not exclusive to and I don't want you to focus
16 on this, but one of the points that caught my attention were,
17 for example, a journalist or somebody else feeding some text or
18 some information into chat explorer for readability. Is this
19 somebody, then, who is actually putting in a draft article and
20 to try to make it -- you know, or is it something else? I
21 mean, I don't know, but I want to understand a little bit about
22 how the particular uses that The Times uses for ChatGPT and
23 with chat explorer and how they may overlap or not with any of
24 the privileges and how they are relevant to those specific
25 various factors. Okay?

P8CEOPE2

1 That was way too long. The written order will be a
2 little bit more clear. But I am concerned about relevance,
3 proportionality, and the scope of the privileges.

4 Ms. Brook, you look like you have a question.

5 MS. BROOK: All housekeeping.

6 So, is your Honor contemplating simultaneous
7 supplemental briefs on this or --

8 THE COURT: Yes.

9 MS. BROOK: Okay. So, simultaneous supplemental
10 briefs this Friday. With my friends on the other side, may I
11 suggest 5:00 p.m. Eastern on Friday so that we don't keep
12 associates late into the evening.

13 THE COURT: Yes. That is generally one of my requests
14 on Fridays. And I know that I'll be seeing some of you on
15 Friday afternoon, I believe, so --

16 MR. SLAUGHTER: Your Honor, just on Friday.

17 THE COURT: Yes.

18 MR. SLAUGHTER: I was wondering if we could make a
19 request that it be due on Monday. Some of us on our team have
20 a Ninth Circuit argument on Friday in an unrelated matter. So
21 if we could put in those briefs on Monday, even sort of noon on
22 Monday or something like that, it would help just because of
23 the Ninth Circuit. But if you --

24 THE COURT: Well --

25 MR. SLAUGHTER: -- want to keep it then, we'll make it

P8CEOPE2

1 happen, but --

2 THE COURT: Well, we can either do it on that Friday
3 -- I know we have a lot of things on Friday. There's going to
4 be a number of lawyers here for the spoliation retention
5 sampling issue anyway. How about we make it the following
6 Friday.

7 MR. SLAUGHTER: For the brief. I was just talking
8 about when to submit the brief.

9 THE COURT: Supplemental brief. We'll make the
10 following Friday, the 22nd.

11 MR. SLAUGHTER: Thank you, your Honor.

12 THE COURT: Okay.

13 MS. BROOK: And three pages, your Honor?

14 THE COURT: Let's do five. That will give everybody a
15 little bit of breathing room.

16 All right. Moving right along.

17 MS. BROOK: That moves us, your Honor, to docket 380,
18 the first green one.

19 THE COURT: Okay.

20 Ms. Logan, I never want to deny a more junior attorney
21 a chance to speak in court, especially with so many spectators;
22 however this is about scheduling a deposition on the
23 Microsoft's metrics tracking dashboards; right?

24 MS. LOGAN: Yes, correct.

25 THE COURT: Do you want to report to me on where the

P8CEOPE2

1 disconnect is?

2 MS. LOGAN: Yes. I believe the disconnect here is
3 exactly that, and that's the timing of the deposition itself.
4 As I'm sure your Honor recalls, and I won't step through our
5 briefing, but our position has always been that this deposition
6 should happen as soon as possible. When your Honor heard
7 argument on our motion, we walked away from that because we
8 were seeking documentation related to these dashboards to cut
9 through a lot of the discovery issues and the back and forth we
10 had, specifically about what we were seeking, what metrics we
11 were teak seeking, how they would be pulled, et cetera. And
12 after that hearing, your Honor ordered that maybe a 30(b)(6)
13 deposition would help cut through some of this noise, and
14 that's exactly what we sought to take, is a 30(b)(6) deposition
15 of one of Microsoft's deponents, and we wanted to take that as
16 soon as possible, especially considering the upcoming deadlines
17 for serving written discovery and substantial completion
18 deadlines.

19 For news points it just simply does not make sense to
20 stall and have this deposition occur later than it needs to,
21 because we want to get these metrics and get this information
22 in hand.

23 THE COURT: All right. Is there a date that you have
24 proposed?

25 MS. LOGAN: Yes. Well, backing up a little bit, our

P8CEOPE2

1 original deposition notice proposed July 29, 2025. That date
2 has come and passed.

3 THE COURT: Right. When was that first notice served?

4 MS. LOGAN: July 8.

5 THE COURT: All right. Then what happened?

6 MS. LOGAN: We were met with refusal from Microsoft
7 to --

8 THE COURT: Did you have any back and forth about the
9 topics?

10 MS. LOGAN: Yes.

11 THE COURT: Okay.

12 MS. LOGAN: We served a deposition notice curated to
13 these five narrow topics on the dashboard metrics.

14 THE COURT: And that was served on July 8.

15 MS. LOGAN: Correct.

16 THE COURT: So then you had some meeting and
17 conferring on the topics. Did you then serve a new depo notice
18 with modified topics?

19 MS. LOGAN: So, on July 31, we served a -- I wouldn't
20 necessarily call it "new," an updated 30(b)(6) deposition
21 notice for all of the 30(b)(6) deposition topics in the case.
22 We originally served the 30(b)(6) deposition notice with all
23 the topics in the case in March, and this was just an updated
24 notice with some additional topics.

25 THE COURT: Okay. So, the 30(b)(6) topics on the

P8CEOPE2

1 metrics tracking dashboards, those topics have been served on
2 Microsoft since July 8 and they haven't changed since then.

3 MS. LOGAN: Correct.

4 THE COURT: Okay. So you need a date.

5 MS. LOGAN: Yes.

6 THE COURT: You wanted a date by August 15. That
7 isn't going to happen.

8 MS. LOGAN: Yeah. Unfortunately. My time machine is
9 still in the works.

10 But our brief requests a deposition by August 15.
11 That is also fastly approaching, but our stance continues to be
12 this should happen as soon as possible.

13 THE COURT: Okay. Let's hear from Microsoft.

14 When can you get this done?

15 MS. BEYER: Good morning, your Honor.

16 Carrie Beyer on behalf of Microsoft.

17 Microsoft has agreed to prioritize the deposition on
18 dashboard topics, and we intend to propose dates near the
19 beginning of the deposition period, you know, as we're
20 transitioning to them now.

21 The issue here is that, consistent with the purpose of
22 the MDL, that we're trying to do coordinated discovery as
23 efficiently as possible. We want to ensure that the witness
24 who is presented is subject to one deposition on these topics
25 and any other topics that Microsoft designates that witness

P8CEOPE2

1 for --

2 THE COURT: No. No. I already said on this
3 metrics-tracking dashboard, maybe you have a different 30(b)(6)
4 witness on other topics, maybe you have the same witness on
5 other topics later, but let's get this one done. Okay?

6 So, what date have you proposed? And when are
7 depositions supposed to start?

8 MS. BEYER: So, our understanding is depositions are
9 at the beginning of September. We had not proposed a specific
10 date yet because we were hoping to look through the expansive
11 30(b)(6) notice that The Times just served where more than 200
12 topics, including enumerated subparts on the 31st, as well as
13 coordinate with the class, so we had not proposed a particular
14 time yet.

15 THE COURT: Okay. 30(b)(6) deposition on the original
16 narrower topics. The five topics served on July 8 should be
17 done by September 30. How's that?

18 MS. BEYER: I think we can make that work, your Honor.

19 THE COURT: Okay. That's the date. Complete the
20 topics from the July 8 notice by September 30.

21 MS. BROOK: Your Honor, if I just may quickly.

22 The only reason we wanted it in August and not
23 December is there's two critical discovery cutoff dates coming
24 up. There's the deadline of August 19 to issue new requests,
25 and there's a deadline of September 22?

P8CEOPE2

1 Help me out, guys.

2 MS. LOGAN: September 12.

3 MS. BROOK: September 12 for substantial completion.

4 THE COURT: Of document discovery.

5 MS. BROOK: Of document discovery. Correct.

6 And part of my understanding of the Court's utility of
7 these early 30(b)(6) depositions is to allow us time to get the
8 information we need to make good use of those deadlines, and
9 there's always good cause and all of that, but to the extent we
10 could even do it in the first two weeks of September, I think
11 that would be much preferable.

12 THE COURT: Look, I'd like you to try to. I'm giving
13 you September 30 as an outside date, but please don't schedule
14 it for September 29. We will probably have to discuss either
15 those discovery cutoff dates or extend them for specific issues
16 but not, you know, as a whole. I don't want the larger
17 case-wide deadlines dragged along because there's one open
18 issue that's still going on. Okay?

19 I gave you September 30 because, as I hear right now,
20 Microsoft might have some work to do to figure out who their
21 witness should be and properly prepare them. I understand
22 we're in the middle of the summer, and that between back to
23 school, back to college, summer breaks, and summer vacations,
24 that you're going to have to find a witness and prepare them
25 for a deposition to be taken before September 30. But I am not

P8CEOPE2

1 going to put it in mid September. Okay? I don't think that
2 gives anybody enough time. And I don't want to be back here
3 with a request for additional hours of deposition or an
4 additional witness or additional topics because you're
5 complaining that the witness wasn't properly prepared or
6 sufficiently prepared. Okay? So, September 30 it is. If we
7 have to modify those other discovery dates, let's deal with it
8 if it comes up.

9 Okay. The next issue here relates to Copilot and
10 Deucalion. I don't want anybody to stand up yet. We've had a
11 number of discussions about discovery relating to different
12 versions of Copilot or Deucalion. I recall one or two, maybe
13 at the last conference, where we were talking about using a
14 water filter analogy. And sorry to see Mr. Briant is not here
15 today because I think he was dealing with this issue the last
16 time.

17 But to zoom out a little bit, this seems to have
18 dissolved into a debate about what constitutes different, what
19 constitutes a new product versus a new feature, and
20 Judge Stein's ruling about no new products; right? No new
21 products in this case. So, I'd like to understand how and what
22 the big picture dispute is. I'm not going to rule on it. I'm
23 actually questioning whether we would need to have a mini tech
24 tutorial about Copilot and what the different versions are. I
25 don't want to call them new products or new futures, and why

P8CEOPE2

1 they matter, similar to some of the relevance arguments I was
2 hearing earlier about The New York Times' chat explorer logs,
3 and/or whether we also need to talk about that with Deucalion.
4 I feel a little more informed about Deucalion, but I'm not sure
5 that helps as far as understanding the scope of the discovery.

6 So with that, I guess, Ms. Maisel, you were about to
7 start talking, why don't you start. But I don't want to hear,
8 necessarily, argument about which versions of Copilot are new
9 or extensions or evolutions, but I want to understand how each
10 side is thinking about this, and because it's the plaintiff
11 seeking this discovery, how or why it matters. Okay? What's
12 the relevance here?

13 MS. MAISEL: Sure. Jennifer Maisel on behalf of News
14 plaintiffs.

15 So, let's zoom out, I think that is helpful, and it
16 starts with our complaints and the infringement and the
17 problematic activity that we allege in our complaint, which is
18 Copilot -- it's a chatbot product, it has different iterations,
19 there's an enterprise version, there's a consumer version, and
20 these chatbot products, as we allege in our complaints, they
21 regurgitate out News plaintiff's content, in part because they
22 are trained on OpenAI's large language models that were trained
23 on these plaintiffs' content, and they also retrieve from the
24 web, through the Bing Search Index, News plaintiff's content
25 and, again, reproduce that in output or use that copied news

P8CEOPE2

1 article content as part of generating output as part of the
2 chatbot products, as part of search.

3 And these are uses. These are the infringing uses
4 that we allege in our complaint, and these uses are continued.
5 While under the hood there have been changes to Copilot, we
6 don't understand those changes. And I welcome the opportunity,
7 perhaps in a settlement conference-type discussion, to explore
8 what those changes are, because I don't have a good handle on
9 what those under-the-hood changes are at this point. We have
10 no discovery into that.

11 But on the surface, the infringing use is the same,
12 and it's been continuing. Back in our initial motion, filed
13 back on March 31, that was the original Times docket 47, we
14 attached some recent outputs from Copilot, which continue to
15 regurgitate out The New York Times and the Daily News content,
16 again, in the same matter as alleged in the complaint, at least
17 on the surface. I mean, the infringing that we've been
18 alleging all along since day one.

19 At the technology tutorial, we dug a little bit more
20 into the products, and there's a little bit of confusion
21 initially, largely on my part, what about the original Bing
22 chat/Copilot product evolved into in terms of M365. There's
23 another statement in Microsoft's brief distinguishing M365 from
24 a product called Microsoft Copilot Chat. I'm not exactly sure
25 of the differences there, other than that chat product seems to

P8CEOPE2

1 be some subcomponent of M365 chat.

2 Again, I think these issues would be helpful to
3 explore in a settlement conference. But at the end of the day,
4 what we understand, at least from the public-facing material,
5 is that these later versions of Copilot use the GPT-4o, 4o,
6 4 omni model, which has been at issue in these litigations. We
7 have taken extensive discovery into OpenAI's training data for
8 GPT-4o. We conducted a model inspection back in December on
9 the GPT-4o model. We're seeking outputs from OpenAI where they
10 named the GPT-4o. And GPT-4o appears be powering these later
11 version of Copilot.

12 And just to be clear, GPT-4o is in the case for good
13 reason. And GPT-4o was released before the Center for
14 Investigative Reporting, filed its complaint. The CIR
15 complaint has examples of 4o output. So, at least from the
16 OpenAI perspective, GPT-4o has been in the case all along. We
17 understand Microsoft's position and OpenAI agree that 4o is in
18 the case, but we have not, and we can address that. But
19 fundamentally, the same GPT models at issue in the underlying
20 litigation also seem to be powering these later versions of
21 Copilot.

22 And then I understand and I'm sensitive to Microsoft's
23 concerns about there being different custodians, a different
24 architecture, different platform powering these newer versions
25 of Copilot under the hood. And I would appreciate learning

P8CEOPE2

1 more about what those differences are, because it's quite
2 possible that many of those differences are just not relevant
3 insofar as the infringing use is concerned. And if we could
4 really drill down into what those differences are and tailor
5 discovery so that way are only getting what we need for
6 purposes of our complaint and the infringing use.

7 I, likewise, don't want to go through 15 more days of
8 source code review. We're not asking for more custodians. And
9 we just want to get to exactly what we need and tailor
10 discovery, but we just have this fundamental gating issue about
11 whether these versions are in the case or not. And once we get
12 past that gating issue, I think it's a problem-solving exercise
13 of how do we tailor discovery to what we need so that we stay
14 on track and get discovery done well before that February
15 close-of-fact-discovery deadline.

16 THE COURT: All right. Let's hear from Microsoft, I
17 guess.

18 MS. HURST: Good morning, your Honor.

19 Annette Hurst for Microsoft.

20 I think it's Ms. Maisel's comment that on the surface
21 the infringing use in the same that really leads back to your
22 Honor's question of why does this matter. Because if, on the
23 surface, the infringing use is the same and the plaintiffs are
24 able to prove their case with the existing products, then
25 discovery into this new product should make no difference

P8CEOPE2

1 whatsoever in the outcome. It won't change the number of works
2 at issue for statutory damages purposes, it won't change the
3 calculous with respect to other forms of remedies, it won't
4 change anything. All it will do to bring this new Copilot into
5 the case is allow discovery to become a tool of asymmetrical
6 warfare by the plaintiffs.

7 THE COURT: Okay. Let's tone down the language. You
8 had me until we got to asymmetrical warfare.

9 MS. HURST: All right. Well, your Honor, this is a
10 massive expansion. Like, I'm not exaggerating here.

11 THE COURT: All right. So, let's go back to what you
12 are saying, though. If, on the surface, the infringing use is
13 the same and -- this is the flip of the chat explorer argument,
14 right -- and if there's an admission that these all rely on
15 GPT-4o, what happens to all the newer versions of Copilot if
16 it's found that they relied on -- I mean, maybe it doesn't
17 matter. Maybe it is just about damages and about what changes
18 have to be made. I mean, I'm trying to understand
19 hypothetically, and maybe this is a question for Ms. Maisel.

20 MS. HURST: I think it's a question for Ms. Maisel,
21 your Honor.

22 THE COURT: Yeah.

23 MS. HURST: Because from our perspective, this product
24 isn't fused in the complaint, it didn't exist at the time the
25 lawsuit was filed, and it is so different, from a

P8CEOPE2

1 proportionality perspective, that it really would be a dramatic
2 problem for us in discovery. We're racing to finish our queue
3 of 150,000 documents for that substantial completion deadline,
4 which was based on the discovery that was served in the spring.
5 We are racing to finish that substantial completion deadline
6 for that discovery. To add this now would just completely blow
7 all the deadlines, your Honor. It just absolutely would.

8 THE COURT: But this is what I'm trying to get to. I
9 hear that, I heard that, I saw it in the papers. But what I
10 also heard from The New York Times is that if -- I'm going to
11 use other analogy, and bear with me because it's not a great
12 analogy. But if the dispute is about a gas-powered combustion
13 engine and the defendant is a car maker and they're using the
14 same engine to power each car, I would agree that the discovery
15 should be really focused on the engine and not the entire car.
16 The car could be very different, it could look different, it
17 could drive differently, it could function differently, but the
18 engine is the accused part. Right?

19 And if that's the case and the plaintiffs win on the
20 engine infringing in some way, I think what I heard -- and
21 maybe Ms. Maisel can be clearer -- is, is it really the issue
22 that we're talking about the engine, 40, or is there something
23 else related to that that may relate to the claims in this
24 case? And I don't know how we get there. You know, because
25 the information is uniquely in the possession of Microsoft

P8CEOPE2

1 right now.

2 MS. HURST: Your Honor, let me suggest this. To
3 extend your analogy, I'm going to go back to the one I used at
4 the technology tutorial, which was the Ford mustang. Right?
5 Same name, different under the hood.

6 THE COURT: Uh-huh.

7 MS. HURST: But if the claim is really whether the
8 name "Mustang" can be used or not, then what's under the hood
9 isn't going to make any difference. And, your Honor, I would
10 submit that if we're just talking about real politic here,
11 that's the answer.

12 You know, if the plaintiffs can succeed in proving
13 their case, then it's not going to really matter. And I would
14 submit, your Honor, if it's fair use then it's not really going
15 to matter, because they're going to be big-picture decisions by
16 the Court which Judge Stein has indicated he's eager to get to
17 and resolve. And, you know, having to do additional discovery
18 with hundreds of new custodians and noncustodial sources is not
19 going to, at the end of the day, affect that, but it will
20 affect the timing of the resolution of this lawsuit in material
21 ways.

22 THE COURT: So, to use your analogy, Ms. Hurst, you're
23 saying it's like the Ford Mustang, it's the same name but is
24 different under the hood, but what I'm hearing from The New
25 York Times is that it's not completely under the hood because

P8CEOPE2

1 it relies on GPT-4o, maybe that's the disconnect? I don't
2 know. Let me ask Ms. Maisel.

3 MS. HURST: Yeah. And I will just say, your Honor,
4 the only reason 4o is -- it was not in the pleadings. The only
5 reason 4o is in the case is because OpenAI answered an
6 interrogatory, number 11, that was directed to OpenAI and
7 agreed to provide discovery about its 4o model on its behalf.
8 Microsoft never agreed to that, and in fact, consistently
9 maintained that it would limit the scope of its discovery
10 related to 4o only to the things that directly involved OpenAI.

11 So, for example, OpenAI's API for 4o runs on Azure and
12 there's revenue sharing related to that, and Microsoft agreed
13 to produce all that consistent with OpenAI's agreement, but
14 Microsoft never agreed that 4o was broadly in the case as to
15 all of Microsoft products.

16 And, your Honor, let me say what else is at stake
17 here. It's not just this new Copilot issue. They're using 4o
18 to go after all of the integration into the office suite of
19 products, which were not accused in the complaint and have
20 never been at issue. And that is why we have consistently said
21 4o is not in the case broadly as to Microsoft.

22 THE COURT: Okay. Let's hear from Ms. Maisel.

23 MS. MAISEL: Let me first address your question, your
24 Honor, and then I'll respond to a couple points Ms. Hurst made.

25 Just the big-picture why does this matter, the

P8CEOPE2

1 dashboard and metric dispute is a helpful exercise here, just
2 as one example. Right now, we do not have -- Microsoft has
3 said that the metrics about the usage of the earlier versions
4 of Copilot are no longer as sensible. We have to undertake the
5 burdensome sampling process of the logs and derive information
6 from the logs.

7 By contrast, for the newer versions of Copilot, those
8 metrics are still live. We can readily query those dashboards,
9 at least from what I understand from Microsoft's documents that
10 it can produce, to ask questions like, how is Copilot being
11 used? Is it being used in substitutive ways?

12 And this analysis it is incredibly important, both for
13 our damages case and for fair use factor four when we're
14 considering the impact on the market. Our complaints allege
15 ongoing infringement. We're seeing ongoing infringement with
16 the newer versions of Copilot. And practically speaking, when
17 it comes to summary judgment and fair use, we're going to be in
18 a situation where we're not going to have the discovery into
19 the substitute of uses of these newer versions of Copilot if
20 they're not in the case. We're not going to see how users are
21 using those products.

22 And right now, again, I don't know differences under
23 the hood between new Copilot and old Copilot, different
24 versions of Copilot, with respect to any mitigation measures
25 that Microsoft may have implemented. We just don't know that

P8CEOPE2

1 right now. And I'm concerned we're going to see arguments from
2 the defendants that things have changed, and we're not going to
3 have discovery to rebut those points or perhaps make our own
4 points about the substitute of uses of Copilot.

5 So, there are very real practical effects of setting
6 this date limit of cutting off our discovery into Copilot and
7 all of its uses, beginning October 1, 2024, which is when one
8 of these newer versions of Copilot was launched. So it's a
9 very practical impact. And under factor four, as
10 Judge Chhabria made clear in this recent decision in the
11 *Kadrey v. Meta* case, factor four is not stuck in time, it's
12 forward looking. We have to look at the damage and the
13 substitute of uses today, and also the likelihood of those uses
14 in the future if the infringement were to continue, become
15 widespread, and be unchecked.

16 So, I think that we can tailor discovery to these
17 issues, for example, into the metrics and dashboards. That
18 would be one category of discovery that would be pertinent to
19 the fair use analysis, but we'd want to see evidence of other
20 substitutive uses as well as the *prima facie* infringement
21 evidence that we've been trying to get.

22 So, again, I think there is a problem solving exercise
23 here on how do we tailor discovery to really get at the core
24 issues in the case and maybe side step some of these issues and
25 the burden on Microsoft in collecting source code and

P8CEOPE2

1 collecting all sorts of documentation about a product that
2 just, quite frankly, may not be relevant to these key issues.

3 THE COURT: Anybody have anything else they want to
4 add?

5 MS. HURST: Just briefly, your Honor.

6 It's not possible to tailor discovery to answer the
7 question how is this product different because it was a
8 start-from-scratch, completely different product, and so it
9 would not be a tailored discovery exercise to inquire into
10 that.

11 With respect to Ms. Maisel's point about we don't want
12 to hear they're going to say it's different, the only way I can
13 possibly imagine that coming up would be if the plaintiffs had
14 prevailed in the case and were seeking a motion for permanent
15 injunction. And we are so far away from that scenario in terms
16 of then litigating a workaround question.

17 Your Honor, even if we were in that scenario, that
18 should only be litigated with the benefit of whatever the Court
19 ruling is at that time about what does and doesn't constitute a
20 fair use. If that became necessary, then we could have a
21 period of discovery at that time that would be directed to the
22 remedial phase of the case. But right now, to do that would
23 derail our schedule.

24 MS. MAISEL: So, just one quick point.

25 I mean, the alternative for us is to file a new

P8CEOPE2

1 complaint with these newer versions of Copilot, and because of
2 the overlapping issues, the same models at issue with OpenAI,
3 it seems like vast waste of judicial resources to do that and
4 to start over from scratch. So, again, that is the
5 alternative.

6 And finally, just before I forget, I want to address
7 your Honor's question about Deucalion. That is a separate
8 issue. While it was raised in connection with the same motion,
9 Deucalion concerns the versions of Copilot that are in the
10 case, there's no dispute about that, the earlier versions of
11 Copilot.

12 If your Honor has any questions about Deucalion, my
13 colleague Kristen Logan can answer those questions, but
14 otherwise we don't see a reason to defer ruling on Deucalion
15 while we work out the Copilot issues.

16 THE COURT: Before you switch to Deucalion, I'm going
17 to direct you meet and confer. Because what I'm hearing from
18 the plaintiffs is a lot of the differences may not matter, but
19 are there specific narrowly tailored things that don't involve
20 restarting discovery, because I also don't have any interest in
21 doing that, that can allow you all to table the issues or put
22 this aside in some way so that there's not a filing of a new
23 complaint that may or may not have to be stayed pending this
24 anyway, and where Microsoft's concerns are also heard?
25 Meaning, maybe there is a way to stage some of this or find a

P8CEOPE2

1 way to defer some of this while keeping the rest of our
2 deadlines intact.

3 Meet and confer, because it seems to me that what
4 plaintiffs are saying is they're not going to ask or they may
5 be willing to settle for something significantly less than what
6 they're asking for right now, but because they're getting the
7 response of, "You don't get anything," you're at these two far
8 ends, and maybe there's a little bit something that you can get
9 that will allow us to defer this issue unless and until, as
10 Ms. Hurst was saying, there's some more findings on the merits
11 that would necessitate some of that, whether we're talking
12 about a new junction or we're talking about scope of damages.
13 Let's see if we can put some of that out; okay?

14 MS. MAISEL: Thank you, your Honor.

15 THE COURT: Okay. And then Deucalion. I want to just
16 quickly get some information on Deucalion and then we're going
17 to take break.

18 So, I agree Deucalion is a separate issue. So, go
19 ahead, Ms. Logan.

20 MS. LOGAN: Thank you, your Honor.

21 I'm going to start by addressing your question at the
22 outset, and that was how is the scope of discovery impacted
23 here. I won't rehash everything we discussed at the first
24 hearing on this motion unless your Honor has particular
25 questions.

P8CEOPE2

1 But just to set the table here, for Deucalion, there's
2 no dispute between the parties that this model is relevant and
3 in the case. Microsoft's argument is a burden and
4 proportionality argument. And during our last hearing --

5 THE COURT: And remind me again what plaintiffs want
6 with respect to Deucalion.

7 MS. LOGAN: Yes.

8 THE COURT: You want to be able to test and run --

9 MS. LOGAN: Yes. Similar to the testing we did for
10 OpenAI's models through its API, we want to do the same with
11 Deucalion.

12 And here, as we discussed at the first hearing on this
13 issue, we still don't understand the specifics of Microsoft's
14 alleged burden here. We don't know how many human hours that's
15 going to take, how many dollars it would take to rebuild what
16 they say, they say they have to rebuild the infrastructure in
17 order to get Deucalion up and ready for us to test.

18 And I think the second point here is most of this
19 burden is self-inflicted. The infrastructure for Deucalion was
20 torn down after News plaintiffs filed suit against Microsoft.
21 So, to the extent that Microsoft now says they have a burden
22 for rebuilding this infrastructure, it was in place at the time
23 of filing.

24 And those are the two points that I have. I'm more
25 than happy to answer any additional questions that you may

P8CEOPE2

1 have. But being mindful of the time, I can --

2 THE COURT: Let's hear from Microsoft on the burden,
3 because this is another proportionality argument. Are we
4 keeping --

5 MS. HURST: Annette Hurst for Microsoft, your Honor.

6 So, the burden here is related to the fact that the
7 product was deprecated, so it no longer exists. And while it's
8 true that that happened during the pendency of the lawsuit, I
9 don't believe the lawsuit, absent some express order, freezes
10 the defendants' products as of the time of the filing. They've
11 continued to develop.

12 So, in order to reconstitute -- I don't even know if
13 it's possible to reconstitute it, your Honor, but it would take
14 a lot of human engineering effort to do so, and there are
15 alternatives.

16 I mean, the first alternative is that plaintiffs had
17 free access to the product using this model for many months and
18 could have inspected and performed whatever operations on it
19 during that period of time. We don't know whether they did or
20 didn't. You know, we'll see that in their expert reports.

21 THE COURT: So, you're saying before it was
22 deprecated, before the filing of the lawsuit, plaintiffs had
23 free access --

24 MS. HURST: Before and after the filing. For a long
25 time after the filing of the lawsuit. This was not deprecated

P8CEOPE2

1 until October 1, 2024. So they had more than a year of free
2 access to this product to test it however they wanted to.

3 But in addition to that, your Honor, the user logs
4 that we're still working on will have data about which model
5 was used. And so when the user log sampling is performed, they
6 can do a specific analysis of how that model was performed
7 through the user output logs, your Honor. And so there was
8 free access to it at one time, there is now a substitute for
9 trying to reconstitute it.

10 You know, your Honor, just with the 80,000 chat logs,
11 we don't know exactly how many lawyers and how much time that
12 is going to take in burden for The New York Times. We don't
13 know how many engineers and how much time it's going to take to
14 put a product back together again. But this is very much a
15 Humpty Dumpty situation, your Honor. It would be extremely
16 difficult and costly to try to reconstitute it, and we do have
17 alternative sources of data.

18 THE COURT: Ms. Logan.

19 MS. LOGAN: Yes. I would like to respond to
20 Ms. Hurst's point about the output log data, and I would like
21 to refute that that would be a sufficient source for this
22 evidence. I would like to remind your Honor, as we discussed
23 at the previous hearing, we are alleging infringement at a few
24 different steps in the process: Infringement during training
25 that the model itself was infringing, and that there are

P8CEOPE2

1 infringing outputs coming from the model.

2 THE COURT: I'm sorry. When you say "during
3 training," do you mean during ingestion?

4 MS. LOGAN: Yes. Ingestion would be another --

5 THE COURT: Okay.

6 MS. LOGAN: But here, the testing that we're seeking
7 to do gets at that second form of infringement, and that's the
8 model itself as an infringement. And that's not something that
9 we can glean from the user output logs that we are planning to
10 get later in the case.

11 THE COURT: So, if your issue was with the model
12 itself, isn't that 4o? Isn't that GPT-4o?

13 MS. LOGAN: So, Deucallion here, it started with --
14 and I believe it's GPT-3.5 Turbo. And Microsoft did further
15 fine-tuning on top of that, and that fine-tuning influences and
16 affects the behavior of the model. And so in order to get at
17 this particular model, we would have to test the version that
18 Microsoft has fine-tuned.

19 THE COURT: And you're not able to get or you don't
20 have discovery on how the 3.5 Turbo version of GPT worked?

21 MS. LOGAN: Yes. In our model instruction with OpenAI
22 we did test the 3.5 Turbo model, but this would be a separate
23 model because they have done further fine-tuning on top of it,
24 so the results that we get from testing the 3.5 Turbo model
25 might not be able to translate and might not be able to apply

P8CEOPE2

1 those to Deucalion as well, because there is further training
2 and fine-tuning on that model.

3 THE COURT: Have you been able to get discovery on the
4 fine-tuning that Microsoft did?

5 MS. LOGAN: Yes. We have access to the fine-tuning
6 datasets for Deucalion.

7 THE COURT: And so why isn't that enough?

8 MS. LOGAN: Because it still doesn't get at the
9 question of what is the model's behavior when prompted. We can
10 see how and what was used to fine-tune it, but that's a
11 different question than what does the model output like when
12 prompted, and what actually is retained by the model, and what
13 is memorized within the model.

14 And for the record, the list of infringing points I
15 provided is not exhaustive, it's merely illustrative.

16 THE COURT: Anything to add, Ms. Hurst?

17 MS. HURST: Just briefly, your Honor.

18 You know, in terms of the output logs being an
19 adequate source of the information, I believe a number of the
20 sampling issues that the Court will be discussing between
21 plaintiffs and OpenAI are about how many output logs are
22 necessary to measure the very thing that we're talking about.
23 Even though the plaintiffs had model inspection of all of
24 OpenAI's models that OpenAI was able to provide, they're still
25 coming back to the Court and asking for the model logs, the

P8CEOPE2

1 output logs, to demonstrate exactly the same information. So
2 that tells us that they view the output logs as adequate
3 information about the performance of the models.

4 THE COURT: Well, the problem with the OpenAI output
5 logs is that it was framed as a spoliation issue, whereas I
6 think Microsoft's output logs didn't have that issue. Right?
7 And you have a lot more of them. So, to some extent, it's a
8 little bit like comparing apples and oranges because the volume
9 of output logs that were lost or destroyed by OpenAI covers a
10 period in time that now we're trying to look back and do a
11 comparison to assess the prejudice of whether losing that
12 volume of output log matters. So, I don't think it should be
13 construed as an admission by the plaintiffs that output logs
14 are all that matters, the issue came up because of the
15 destruction of the output logs.

16 What I am trying to get at here is whether the
17 inability to -- because Microsoft is refusing, at this point,
18 to make Deucalion -- to reconstitute Deucalion so that
19 plaintiffs can run their own searches, I'm trying to understand
20 what the adequate substitute is. So let's get them all out and
21 you'll remind me what I'm missing. So I've heard, first, that
22 plaintiffs were able to run searches for months before and
23 after.

24 MS. HURST: Yes.

25 THE COURT: Right?

P8CEOPE2

1 MS. HURST: They have the training data. And they
2 will have the output log data, which can give them an adequate
3 sample of the performance of the model log.

4 THE COURT: Right. And Microsoft's output log data
5 actually dates from the time that Deucalion was active.

6 MS. HURST: Yes, your Honor.

7 THE COURT: Up and running.

8 MS. HURST: Right.

9 THE COURT: So we don't have the data loss issue that
10 we have with OpenAI.

11 Okay. And then that's it. Those three categories;
12 right? Or buckets of information: The searches that
13 plaintiffs were able to run before and after the filing of the
14 lawsuit before Deucalion was deprecated, the training data
15 which has been produced, and the output log data which is
16 contemporaneous and was going to be produced.

17 MS. HURST: Correct.

18 THE COURT: Okay. Anything else?

19 MS. HURST: No, your Honor.

20 THE COURT: Okay. So then the question for plaintiffs
21 is, I mean, this does sort of shade into the proportionality
22 argument, even though I don't have super clear information on
23 the burden of re-creating Deucalion, and we're not going to
24 have the technical knowledge to understand what that takes, but
25 why are these three categories of information, once you get

P8CEOPE2

1 them all -- what is still missing and how is that proportional
2 to what you're already getting produced in the case and all the
3 proportionality considerations of the Rule 26?

4 MS. LOGAN: Sure. So, just as an example, this
5 testing would show how much the model has memorized. And
6 neither of these sources of information would show that. For
7 example, the training data, I'm sure defendants would not
8 accept that because there is all this training data, all that's
9 memorized by the model. We need to test that. We need to test
10 the model to see that.

11 Also, the reference to us being able to test Deucalion
12 before it was deprecated. When you use these tools, it doesn't
13 say: Your query is now being answered by Deucalion. There
14 would have been no systematic way for us to test it through
15 their user interface.

16 And as far as getting the fine-tuning data, the same
17 is true there. That doesn't tell us how much content the model
18 has memorized, nor do the user queries to the chatbot fall
19 illuminate this issue. In order to see that, we do need to do
20 testing.

21 And I just wanted to add, if it is Microsoft's
22 position that they don't want to reconstitute the
23 infrastructure here to run Deucalion, then we would want to
24 reserve the right at trial, then, if Microsoft argues that we
25 haven't shown X, Y, and Z with respect to the Deucalion model,

P8CEOPE2

1 that we would have had it not been for the refusal to
2 reconstitute the infrastructure and allow us to do this
3 testing.

4 (Continued on next page)

P8CJOPE3

1 THE COURT: All right. So I think what I'm hearing is
2 that plaintiff's concern about testing Deucalion has been
3 narrowed a little bit, so we can't see or we don't know how
4 much the model has memorized. Is there another way to get that
5 information that doesn't involve reconstituting Deucalion?

6 MS. HURST: I think the question of how much the model
7 memorized is a "tree fell in the forest and there was nobody
8 there to hear it" question, your Honor. All that matters from
9 the perspective of whether there could possibly be harm to the
10 plaintiffs is was there some substituting output, right? This
11 theoretical question of whether the model in some abstract
12 world has memorized something is not a question about fair use.

13 It's not a question about infringement either because
14 copies were made in the course of the training, and so those
15 are the copies. So none of what the plaintiffs are proposing
16 advances the fundamental considerations and questions in the
17 case when they're talking about some abstract notion of
18 memorization.

19 Let me add that, as we discussed at the tech tutorial,
20 there are a variety of filters on the Copilot product that is
21 the only product that ever made use of Deucalion that would
22 prevent regurgitation. It would prevent copying of data as an
23 output. And so there's no way to look at that anywhere. And
24 it didn't exist before, and it didn't exist now.

25 But to the extent there are outputs that are some kind

P8CJOPE3

1 of a regurgitation, those will be in the output logs. And I
2 think that when they say "memorization," what they really mean
3 is regurgitation. That's the only thing that would possibly
4 matter, even assuming that it does, and that would be reflected
5 in the logs, your Honor.

6 THE COURT: Ms. Logan, any clarification on that?

7 MS. LOGAN: Just a couple more points, your Honor. To
8 respond to Ms. Hurst's "tree fell in the forest" point, I would
9 like to argue that this is directly relevant to factor one and
10 transformativeness as well. This is a key issue in the case,
11 and what is actually happening here during training and how are
12 those works retained in the model.

13 And second, OpenAI distributed its models to
14 Microsoft, and that's a separate act of infringement. And we
15 reserve the right to test that model that had been distributed.
16 Thanks.

17 THE COURT: Okay. Let's take a break 15 minutes.

18 (Recess)

19 THE DEPUTY CLERK: Court is back in session.

20 THE COURT: Before I tell you all what I'm going to do
21 as far as Deucalion, I wanted to get a little bit more clarity
22 on what everybody means by "memorization." What stage, and
23 what is -- what do plaintiffs mean by "memorization" and what
24 do defendants mean by "memorization"? And are we even talking
25 about the same thing? Sorry, Ms. Logan.

P8CJOPE3

1 MS. LOGAN: No worries. Always happy to talk about
2 memorization.

3 By memorization, we – and did I hate to use the word
4 to describe it – but we mean the training data that has been
5 memorized by the model. And that's the data that is encoded in
6 the model's parameters. So for example, out of the training
7 data, X, Y, and Z articles might be memorized by the model.
8 It's encoded within the model's parameters. And one way of
9 getting at this and one way of empirical testing to figure out
10 whether it's been memorized is to test for regurgitation. But
11 regurgitation is not synonymous with memorization.
12 Regurgitation is just how much of this stored training data can
13 we actually get the model to output, but it is a separate
14 concept technically.

15 THE COURT: But you have the training data and the
16 data sets?

17 MS. LOGAN: Correct. But the inquiry is a bit
18 different. Even if we have the training data, we still want to
19 figure out how much of that has in fact been memorized by the
20 model.

21 THE COURT: So let's step away from the word
22 "memorized" because it's very anthropomorphic. So as I
23 understood it, if there is training data that goes into the
24 model, it's broken up into bits and it's analyzed -- in a way
25 it's groups of data that occur near each other, and there's

P8CJOPE3

1 different ways of weighting that data and that analysis, right?
2 So how is that analysis and the information there, right, the
3 probabilistic associations of data, whether it be words or
4 quote/unquote concepts, which is also a little anthropomorphic,
5 how is that not retained? Of course it's retained in the
6 model. So how is memorization different, if at all?

7 MS. LOGAN: Sure. Yes. And we agree that is retained
8 in the model. But we want to do testing to get at that because
9 defendants are going to argue that these things are -- it's not
10 memorizing the actual articles, it's just memorizing
11 associations between the words. But those associations between
12 the words constitute the entire articles. And so we want to
13 get the evidence that we need to support those points --

14 THE COURT: But I mean, isn't this -- this is why I
15 wanted to get away from the word -- from the expression or the
16 concept of memorization. Because if there's data that's fed
17 into it, it's words that are close together, right? Series of
18 words, and they occur within certain distances from each other.
19 And then assume that there are some weight or something to the
20 quality of what's been input so that it weights certain
21 associations or certain patterns of associations heavier than
22 others. That is part of the model, so how is memorization
23 different from that?

24 MS. LOGAN: Yes. And I believe here we want to get
25 evidence that that has in fact occurred, correct. As you

P8CJOPE3

1 recall from the tech tutorial, we presented evidence and we
2 discussed how certain high quality content is up-weighted. And
3 that might result in this content being encoded in the model's
4 parameters. This is further evidence to show that.

5 We want to be able to support this argument by showing
6 – and I will refrain from using the word "memorize" – by
7 showing that the model can in fact spit out those exact works
8 that it was trained on. And this goes directly to factor one
9 and the transformativeness prong.

10 THE COURT: Okay. Can I hear from either Microsoft or
11 OpenAI on whether either I or plaintiffs are mistaken or
12 misunderstanding.

13 MS. HURST: Your Honor, I'll just use the word
14 "memorization" two more times including that one.

15 We would characterize it thusly – memorization is the
16 capacity to regurgitate. So there's not a meaningful
17 difference. And the only way it can be directly -- it can't
18 really be directly observed because, as has been the subject of
19 considerable scientific study, people don't have a full
20 understanding of how exactly these models work, but to the
21 extent that they can be induced to create an output that is
22 substantially similar to a piece of training data, the only way
23 that can be observed is through observing the behavior of the
24 model, which is why the plaintiffs have sought model inspection
25 and why the plaintiffs are also seeking a large amount of

P8CJOPE3

1 output data. They are seeking that output data to measure
2 regurgitation. And I absolutely understand, your Honor, that
3 there's a dispute about the deleted data and whether there are
4 differences in measurement as a result of that.

5 THE COURT: That's not Microsoft's issue, right?

6 MS. HURST: No. But your Honor, ultimately the point,
7 what they are trying to measure through output data, whether
8 there are differentiations in the allegedly deleted data or
9 not, is the rate of regurgitation. And so by doing so, what
10 the plaintiffs have implicitly and expressly said is that
11 output logs are an appropriate measurement to observe the
12 performance of the model with respect to whether it has
13 retained a set of associations that allow it to mimic the
14 plaintiffs' data and output.

15 Let me add one other point, your Honor. We have
16 produced -- because we have produced the fine tuning data, the
17 maximum number of works that could possibly be included in the
18 infringement set is known to the plaintiffs. And so they know
19 what the maximum possible -- and it's a very small number, your
20 Honor. It's a very low rate of what's included in the fine
21 tuning data. I won't go into the particular numbers on the
22 record, but it's a tiny number.

23 So they are aware that there's a tiny number in the
24 fine tuning data of their material. They can quantify exactly
25 what the maximum rate of that is, and then they can seek to

P8CJOPE3

1 measure whether any of it is referenced in the output data that
2 we will be providing to them. So they have everything they
3 need to paint this picture, paint by the numbers, your Honor.
4 They have everything they need here.

5 THE COURT: Okay. All right. Go ahead, Ms. Maisel.

6 MS. MAISEL: Just very briefly because I think there's
7 a little bit of talking back and forth, and I want to tie this
8 back to the copyright infringement claims and the act of
9 reproduction and distribution.

10 So on the reproduction side, we have the reproduction
11 claims as part of the acquisition and the curation of training
12 data sets. It's part of the ingestion claim.

13 And then we have training. You heard me speak about
14 the loss function at the technical tutorial. This is the
15 actual instructions given during training to help optimize the
16 model so that it's able to accurately predict the next token in
17 its training data, to reproduce its training data. And we are
18 trying to ascertain through these model inspections the extent
19 to which that training process, that loss function and the
20 other -- and the fine tuning and the midtraining and all those
21 training processes, to what extent there actually is or is not
22 transformation of that underlying training data and to what
23 extent, if there is -- one end of the spectrum is this
24 phenomenon memorization where the model is simply encoding
25 copies of the training data within its parameters such that it

P8CJOPE3

1 can be regurgitated out.

2 At the other end of the spectrum, which we heard a
3 little bit at the tech tutorial, is no, we're just extracting
4 facts, statistics, other information about language so that we
5 can create something new. We heard themes of that at the
6 technical tutorial too, and there's those two spectrums. And
7 the model analysis, the memorization analysis is to understand
8 that transformation, where in the spectrum we are as part of
9 that training process.

10 And I also mentioned distribution because Deucalion is
11 unique in the sense that at least as of today with the older
12 versions of Copilot, it is the only model we're aware of that
13 OpenAI transmitted, distributed to Microsoft, that Microsoft
14 then fine tuned. And I'm not going to get into the fine tuning
15 process. And as Ms. Hurst said, there are news-related cases
16 in that fine tuning data. And we want to know to what extent
17 that fine tuning and that distribution of the model might in
18 and of itself be an infringement, and also to test our
19 hypotheses to what extent and where in that transformation
20 spectrum we fall.

21 That's the purpose of this model inspection. And
22 we've been through this with OpenAI, and we're simply trying to
23 get access to another model, this unique model Deucalion that
24 Microsoft fine tuned.

25 THE COURT: All right. Thank you.

P8CJOPE3

1 So my ruling on Deucalion is that I'm not going to
2 direct Microsoft to reconstitute it at this time. As I hear
3 the arguments today, it is where it is, and if there were a
4 time in the future where it would need to be reconstituted or
5 recreated, it could be done so. So deferring that and putting
6 that off or denying that request as of right now is not likely
7 to lead to the destruction of any information or loss of
8 information. It's just a time delay issue.

9 What I am finding as of right now is that it's not
10 proportional to require Microsoft to recreate or reconstitute
11 Deucalion so that plaintiffs can run tests in light of what
12 Microsoft has put on the record of the information that is
13 being provided. All of this is in this some form or way or
14 another an indirect look into what the model is doing or how it
15 might have been encoded.

16 But because plaintiffs had and were able to run
17 searches on Deucalion when it was up, they've been able to test
18 the model that OpenAI had and presumably can still continue to
19 test on their own if they want. They have the training data
20 and they have the output log data that is in fact
21 contemporaneous with Deucalion.

22 I'm going to rule that as of right now that's
23 sufficient. The output log data is, as I understand, quite
24 voluminous. And if the output log data is only going to be
25 able to see regurgitation or that's what you're trying to see

P8CJOPE3

1 as a sign or as a way in which to infer memorization or
2 encoding of certain training data, as of right now I think
3 that's sufficient. And the jury instructions in this case are
4 going to be a bear is all I'm going to say, and I'm happy I'm
5 not doing them. So that's where we'll leave that.

6 So I think the next issues then are the privilege
7 issues.

8 MS. BROOK: Yes. If I can just play housekeeping for
9 a minute, your Honor, I think we have docket 375 are the
10 privilege logs, docket 383, which are the survey results. It's
11 then on Exhibit C, but the last of the news plaintiffs' issues
12 is exhibit 390, the texts and social media one, which relates
13 to both news and class, and then class has four issues, I
14 believe, on their chart.

15 THE COURT: Okay. So all right.

16 MS. SALOMAN: Good afternoon, your Honor. Sarah
17 Saloman on behalf of defendant OpenAI.

18 This motion concerns two categories of baseless
19 assertions of privilege by the Times. The first is with
20 respect to a training that the Times developed in the leadup to
21 the release of Chat Explorer which my colleague, Ms. Ybarra,
22 explained was the tool by which New York Times employees could
23 use RLLMs. And the second category is presentations regarding
24 demos of RLLMs, and from what little we can glean from the
25 production, concerns about implementation of RLLMs quickly

P8CJOPE3

1 enough that from the document production to date seem like they
2 were created and then disseminated in the key summer and fall
3 2023 period when the times was experimenting with RLLMs.

4 Now, for any privilege claim, the Times needs to
5 establish three things: The first is that the document at
6 issue had a legal purpose and not a business report;
7 the second is that it was confidential; and the third is that
8 the communication or document reflects communications with a
9 legal advisor or communications at the direction of a legal
10 advisor.

11 And I want to be really clear about the documents that
12 are at issue in this motion. Despite five months of meet and
13 confer, the Times has not identified a single privilege log
14 entry that relates to these documents, has not attached a
15 declaration to their opposition to our motion that would
16 indicate anything about the privileged nature or so-called
17 privileged nature of these documents, and, in fact, up until
18 their briefing had not disclosed any facts at all that would
19 purportedly support the privilege.

20 I can go through in detail what we do know from the
21 production and why it suggests that neither of those three
22 elements of privilege have been met.

23 THE COURT: Okay, briefly. And then tell me the
24 relief you're seeking.

25 MS. SALOMAN: Well, I can start with the relief, which

P8CJOPE3

1 is production of these documents which the Times has not
2 substantiated privilege over. With respect to the training,
3 the documents -- and I would direct the Court to exhibits three
4 to seven --

5 THE COURT: No, just talk to me about how many
6 documents are at issue, and then I don't need to look at the
7 documents right now. Because one other way to read this relief
8 requested is that you want an in camera review. And if that's
9 the case, I'm not going to do an in camera review on the bench
10 where we have all these lawyers here.

11 MS. SALOMAN: We don't know how many documents are at
12 issue because the Times has never identified the documents in a
13 privilege log. What has happened is we found references to
14 this training, found references to these presentations in
15 produced documents by the Times, and we said hey, these have
16 not yet been produced. What are they? The response in an
17 email was they're privileged. And we said where are they
18 reflected on your privilege log? We have never received an
19 answer to that question.

20 THE COURT: Let's get an answer to that question.

21 MS. SALOMAN: Sure, your Honor.

22 MR. MUTTALIB: Thank you, your Honor. Adnan Muttalib
23 for the New York Times.

24 It's three total documents, your Honor. There's the
25 training module and there's the two presentations at issue --

P8CJOPE3

1 THE COURT: Three buckets or three documents?

2 MR. MUTTALIB: Three total documents. First, is the
3 generative AI training module. That's the first category. And
4 the second category are the presentations. And from our
5 reading of OpenAI's motion, they're challenging the assertion
6 of privilege over two of those presentations.

7 THE COURT: Other presentations have been produced?

8 MR. MUTTALIB: We have produced hundreds of
9 presentations related to generative AI. We have only withheld
10 presentations that reflect legal advice from the Times in-house
11 counsel relating to this litigation, potential litigation, or
12 investigation regarding generative AI products for the purpose
13 of this litigation. And we believe these two presentations
14 fall in that category.

15 THE COURT: What about the GenAI training module?
16 That's two presentations. I can review that in camera.

17 MR. MUTTALIB: Of course, your Honor, and we'll
18 produce those documents to you if you like. The generative AI
19 training module, I think the best way to go about it is OpenAI
20 raises two fundamental questions. First is the training
21 privilege; second is the distribution of the training breaking
22 that privilege.

23 THE COURT: What? Training privilege, and then what's
24 the other thing?

25 MR. MUTTALIB: Does the dissemination of that training

P8CJOPE3

1 break the privilege. In terms of whether the training is
2 privileged, the first, I think, matter is the training was
3 prepared to ensure that Times employees were utilizing
4 generative AI tools in a legally compliant manner and to inform
5 them of the risks of failing to do so.

6 So the purpose is very clearly legal, contrary to what
7 OpenAI is representing. And also contrary to what OpenAI has
8 stated, there were 12 Times attorneys who were responsible for
9 drafting, editing, revising, and finalizing that training. The
10 only basis that OpenAI presented in its briefing is that Shawn
11 Bower, a nonlawyer, also played a role in preparing that
12 training. But Shawn Bower's role is to provide security
13 advice, and it is unsurprising that a legal training that talks
14 about legal compliance also talks about security compliance.
15 But SDNY case law is clear that when a training is prepared
16 with the assistance of counsel, that document still retains its
17 privilege.

18 THE COURT: Okay.

19 MR. MUTTALIB: And if you'd like, I can quickly deal
20 with the dissemination point because they make much of the
21 point that the training was disseminated to hundreds of Times
22 employees. But I think very quickly, the purpose of *Upjohn* is
23 the idea that you can circulate legal advice to a corporation's
24 employees for the purpose of ensuring legal compliance. And I
25 think that's relatively consistent with the position that

P8CJOPE3

1 OpenAI itself has taken with asserting privilege over its own
2 training documents.

3 THE COURT: Okay. And is it true or correct that
4 these three documents do not appear on any privilege logs?

5 MR. MUTTALIB: Your Honor, I believe the generative AI
6 training module is not on the privilege log yet. We have
7 produced 75 percent of our privilege log, but there is still
8 one outstanding privilege log remaining. In terms of the two
9 presentations, I believe for one of them we identified a
10 privilege log ID, and the other one we did not for the same
11 reason.

12 THE COURT: Here's what I'm going to -- we're going to
13 move this along. I'm going to direct the Times to provide a
14 supplemental log specifically for these three documents.
15 You're then going to meet and confer, and then you're going to
16 see if you can resolve your issues. Or if you can't, you're
17 going to let me know and then I will do an in camera review.
18 However, when I do the in camera review, I'm going to make the
19 loser pay the costs of all of this privilege review, okay,
20 everything after the privilege log. So I really want you to
21 have a fulsome meet and confer once the defendants have the
22 benefit of the privilege log and have had some time to digest
23 some of this. All right. Why don't we do this. New York
24 Times, can you do your supplemental log by this Friday?

25 MS. BROOK: Yes, your Honor, of course.

P8CJOPE3

1 THE COURT: Okay. Log provided by 8/15. Meet and
2 confer and file a joint letter on 8/22 whether you've resolved
3 the dispute or would like in camera review of one to three
4 documents with the full knowledge that loser is going to pay,
5 okay?

6 In the 8/22 letter, if you're not able to reach
7 agreement and you want in camera review, you're also going to
8 propose a production and briefing schedule on the privilege
9 issues for that particular document, okay? So you're not going
10 to submit the briefing yet. I expect that whatever schedule
11 you come up with will be fine with me, but I just want to see
12 the schedule and endorse it before I lock myself into meeting a
13 timeline that I can't make, okay?

14 All right. So that takes care of some of the -- are
15 there other privilege related issues that need to be taken care
16 of?

17 MS. BROOK: No, your Honor. That takes care of docket
18 entry 375. Next step is 338, survey results.

19 THE COURT: Oh, yes.

20 MS. NAJEMY: Good afternoon, your Honor. Laura Najemy
21 on behalf of Microsoft.

22 Your Honor, Microsoft is seeking very discrete relief
23 with a motion that is filed here that's before you that the
24 Times be compelled to produce analyses that it created relating
25 to the results of a single survey that it ran related to the

P8CJOPE3

1 reason subscribers canceled their subscriptions in which "AI
2 makes my subscription unnecessary" was an option. This is not
3 an attempt to relitigate any issue that the court has already
4 ruled on. In fact, it's quite the opposite.

5 The Court previously held "that subscription
6 cancellation surveys are relevant insofar as they illuminate
7 whether subscribers canceled their subscription because of
8 generative AI or for some other reason." And the Times itself
9 has that conceded that the survey that it ran in which "AI
10 makes my subscription unnecessary" was an option was relevant,
11 and it agreed to produce that survey. And I believe we also
12 heard from Ms. Brook earlier today that the most important
13 aspect of fair use is whether the product at issue takes away
14 revenue through substitution.

15 What the Times is refusing to produce here is any
16 analysis that it has done of that particular survey where a
17 miniscule number of respondents actually checked the box
18 suggesting that AI had anything to do with their cancellation.
19 It hasn't argued that these analyses do not exist. Instead it
20 has argued that it has produced some materials that contain
21 some analysis relating to surveys, but it will not search for
22 or produce any others. But that's now how discovery works.
23 The Times doesn't get to cherry-pick here what relevant and
24 responsive documents it wants to produce or it would prefer to
25 produce, as opposed to others.

P8CJOPE3

1 And what's worse is that the Times hasn't produced any
2 of these audience insight reports that it points to that would
3 cover the timeframe in which the particular survey was run. It
4 was run from July 19, 2024 to January 2, 2025, and the audience
5 insight reports that it has produced to date end in October of
6 2023.

7 To the extent that it has produced such slide decks
8 that make reference to other survey results which would be
9 earlier than the survey that we are concerned with here, those
10 reports or presentations contain at most three to four pages
11 discussing the survey results and are largely just charts that
12 provide the information without substantive discussion or
13 analysis of what those results actually mean.

14 And I would also note that to the extent there is any
15 substantive discussion in these presentations, it revolves
16 around the major reasons for cancellation, not those that just
17 show up in a handful or less of responses for the survey. And
18 the document that the Times attached to its opposition doesn't
19 even discuss any cancellation survey at all. It seems to be
20 the relevance there is that it discusses the effects of other
21 outside factors that may have impacts on the Times's
22 performance, but it has no discussion in 167 pages of any
23 cancellation survey.

24 So in short, it is undisputed that this survey, this
25 single survey that the Times ran, is relevant, and the Times's

P8CJOPE3

1 own analysis of that survey of how it understands and
2 interprets these results is equally relevant here to the
3 Times's claims that defendants' products are causing it harm.
4 How the Times itself has dissected those results, whether it
5 found that AI is truly causing harm or not causing harm,
6 whether they concede in their own analysis based on the fact
7 that such a tiny number of people selected AI as a reason
8 really shows the defendants' products are not causing any harm
9 at all is relevant to the case and must be produced.

10 This is a simple go and get. The Times can easily
11 identify who is responsible for the survey and ask them to
12 provide any analyses or reports that they have created relating
13 to this particular survey. It's not burdensome, and this is
14 exceedingly relevant to the central issue in this case, as the
15 Times put it today, that whether or not the product at issue is
16 taking away revenue through substitution. So any
17 proportionality argument that the Times comes up with today,
18 which I will note was not in the briefing that they filed or in
19 their opposition, simply can't outweigh the very clear
20 relevance of these analyses.

21 THE COURT: All right. Let's hear from the Times.

22 MR. MUTTALIB: Your Honor, you must be sick and tired
23 of me coming up here and talking about surveys because this is
24 the third time that we've done this. On November 18, 2024,
25 Microsoft filed a motion to compel subscription cancellation

P8CJOPE3

1 information. That included reasons for any cancellation of
2 subscriptions to the Times, surveys and responses regarding
3 cancellation of subscriptions, and studies regarding the
4 reasons for any cancellation of subscriptions to the Times.
5 That clearly covers what they're seeking here – namely,
6 documents related to subscription cancellation information and
7 documents related to survey data. The Court denied that
8 request at docket 344, and Judge Stein affirmed that order at
9 docket 592.

10 At the last hearing when we were arguing about the
11 Times's trademark claim, Ms. Hurst stood up and said this is
12 why we need surveys. Your Honor once again denied that
13 request. So for a third time we are now litigating this issue
14 again of whether they're entitled to additional documents on
15 top of the survey that we already produced. And I think the
16 answer very clearly based on this Court's prior ruling and
17 Judge Stein's affirmance of that order, is no, that survey is
18 sufficient.

19 And I think there's good reason for that because we've
20 produced swaths of data relating to this exact issue. So on
21 that basis alone we think Microsoft should be precluded from
22 raising this motion again. But to the extent your Honor would
23 just like a recap of what we've produced, let me be clear about
24 what that survey entails, because I think their motion is
25 silent about what is in that survey.

P8CJOPE3

1 That survey is not just surveyor number one, reasoning
2 for cancellation. That survey includes extensive detail for
3 over 15,500 survey responses that includes the following
4 information about subscribers: Their age, their geography,
5 their gender, their employment status, their income, their
6 political ideology. Those are just some of the facts that are
7 there about the subscribers.

8 On top of that, it includes information regarding why
9 the subscriber initially subscribed to the New York Times, why
10 that individual unsubscribed, and the overall impressions that
11 individual has of the New York Times relating to political
12 reporting, bias, all the things that they allegedly need this
13 additional reports and analysis for. They haven't
14 substantiated why that survey that has over 15,500 responses is
15 insufficient for the type of data they would need to understand
16 the survey responses.

17 Furthermore, we've produced audience reports for the
18 very date range that Microsoft is now seeking these surveys,
19 and we've produced the audience insight group analyses, which
20 are the Times's own analysis about the impacts of generative
21 AI, including defendants' products on the Times. And when you
22 read their motion, this is exactly what they're seeking through
23 this redundant request again, which is why this Court has
24 consistently denied Microsoft's request for survey data.

25 Your Honor, I think the final point that I want to

P8CJOPE3

1 make is -- and I say this not to -- just as a matter of fact --
2 is this is kind of how this keeps happening. We get a request
3 for a subset of documents. We're told that it's a narrow
4 request, this or the cancellation information, we produce that
5 document and then we get a followup request saying give us
6 every document, analysis, and report related to that document
7 you produced.

8 At some point this has to stop. We put a lot of work
9 into producing that survey, and now to have to go and produce
10 every report and analysis related to that survey is excessively
11 burdensome, and I think is unnecessary given what they already
12 have.

13 I'm happy to answer any questions, but I think for the
14 reason that the Court has already ruled prior in its prior
15 order and at the June conference, this Court should deny
16 Microsoft's request.

17 THE COURT: My question for both sides is why isn't
18 this -- or isn't this an expert's issue? Mr. Muttalib, you're
19 standing up.

20 MR. MUTTALIB: I think you're absolutely right, your
21 Honor. I think first of all, this can be better understood at
22 depositions. If they have questions about the survey, they can
23 ask questions about the surveys at the relevant depositions.
24 But second, to the extent this gets into technical details
25 about cancellations and analyses about that, of course it would

P8CJOPE3

1 be better reserved for expert discovery. So I think that's
2 absolutely right and should be reserved at least until that
3 time, to the extent the Court is even inclined to consider this
4 argument. And to the extent your Honor – and you actually made
5 this point at the last hearing – to the extent they want to do
6 surveys, they can do that with their own experts at that time.

7 MS. NAJEMY: Your Honor, no expert in this case can
8 substitute for the New York Times's understanding of its own
9 business model. The fact that there is a huge amount of data
10 that they have produced underscores the need for their
11 analysis --

12 THE COURT: Right. So their analysis is going to show
13 up probably in the best forum in an expert report, right?
14 Because that's what they're going to be held to at trial.

15 MS. NAJEMY: No, your Honor, not the Times's own
16 understanding of its own business model and how these
17 particular surveys -- we're talking about a survey over 15,000
18 people. And those survey results – which I understand have
19 been marked attorneys' eyes only and I cannot disclose – show
20 that a very miniscule number of people who are responding to
21 that survey said that AI had any impact on them canceling.

22 To the extent that the Times is having discussions
23 internally about what those survey results mean, it directly
24 impacts the harm that they are claiming here. Only they are in
25 the best position to understand how these survey results are

P8CJOPE3

1 impacting their business. if they look at these and they say X
2 number of people canceled and said it was solely based on AI
3 and that really doesn't impact our bottom line, that's
4 important to this case. But what they are refusing to do is
5 find what I cannot imagine are particularly large volumes of
6 analyses, given that this is a single survey that was run from,
7 again, July of 2024 to January of 2025, it is simply asking the
8 people who are in charge of running that survey, did you
9 produce a report or an analysis that looked at the data that
10 you were looking at and you were testing from your subscribers?

11 And I also point out that Mr. Muttalib said they
12 produced these audience insight reports. That is absolutely
13 false. They have not produced the audience insight reports
14 from the date range in which this survey was done. And even if
15 they did, from looking at all the audience insight reports that
16 they have produced to date, all we would see is three pages of
17 information that show us charts that just categorize these
18 things and say this is the most common reason, and we are very
19 concerned about the most common reason. But it doesn't tell
20 you about reason number 100 at the bottom of the list and how
21 it's impacting their business, and that is critically relevant
22 to this case.

23 THE COURT: I'm still not getting why you can't just
24 depose witnesses and say, what did you think about the internal
25 analysis? I don't understand what other analysis – it's a meta

P8CJOPE3

1 analysis at that point, right? It's an analysis on the
2 analysis, what you're looking for?

3 MS. NAJEMY: No. It's their contemporaneous
4 understanding of what these survey results mean to their
5 business. Can we ask questions about that in deposition?
6 Absolutely. But the best evidence of how they viewed this at
7 the time they ran the survey is their analysis at the time they
8 ran the survey to the extent that they've documented that. And
9 the Times has made clear that they do have these analyses.
10 They just don't want to go and look for them and get them. And
11 given the importance and the relevance of this issue to the
12 central question of fair use here, there's no proportionality
13 argument that can stand up to how centrally relevant this is.

14 THE COURT: Let's hear the proportionality argument.
15 I'm not necessarily convinced that it's relevant and not
16 duplicative because it seems that, looking at some of this
17 stuff, there might be a very obvious inference right? If you
18 already have the number out of the 15,000 and you're
19 representing that that's very, very miniscule, isn't that a
20 Times a problem?

21 MS. NAJEMY: It is precisely a Times problem, your
22 Honor. But the Times's own word and their own analysis of what
23 that problem is and how little of a problem it actually is is
24 very relevant here.

25 THE COURT: Let's hear from Mr. Muttalib.

P8CJOPE3

1 MR. MUTTALIB: Thank you, your Honor. I just want to
2 respond to three points that Microsoft made. The first is they
3 say they want evidence of the New York Times's own
4 understanding of its business, and I think that we've already
5 made clear that we've produced audience reports dating back to
6 2022, namely the timeframe that they want from the survey
7 responses, which is again the New York Times's own analysis and
8 reports of the reasons for the decrease of traffic to the
9 Times's website.

10 We've also produced AIG reports. And with all
11 respect, I did not misrepresent the nature of our production
12 because those reports did not exist back to 2022, the AIG
13 reports.

14 THE COURT: Okay. I heard — and I may have misheard —
15 that there were no audience insight reports produced after
16 July 2024. Their issue with the time period is that that's
17 July 2024 to January 2025.

18 MR. MUTTALIB: To be clear, your Honor, there's two
19 distinct types of reports, and we should really rename them.
20 But the first is audience reports, which is the Times's
21 analysis of the reasons in decreases for traffic to the Times's
22 website. That has been produced back to 2022.

23 THE COURT: From 2022 to when?

24 MR. MUTTALIB: The end of the date range, which was
25 May 22, 2024.

P8CJOPE3

1 THE COURT: Okay.

2 MR. MUTTALIB: And the very aptly named AIG reports
3 are for the impacts of generative AI on the New York Times, the
4 Times's own impressions of that. That can only be produced for
5 that specific date range because that's when those reports
6 started to be created.

7 THE COURT: So when you say "that specific date
8 range," what do you mean.

9 MR. MUTTALIB: July 2023 onwards.

10 THE COURT: Okay. So those AIG reports, are those the
11 same things that Ms. Najemy -- are those the same things that
12 you're saying are the audience insight reports that were not
13 produced after July 2024?

14 MS. NAJEMY: If I may, your Honor, no reports have
15 been produced after July of 2024, which I think Mr. Muttalib
16 would agree with me on. The problem is that they are telling
17 us that we've produced all of those reports that talk about the
18 survey, but the survey ran from July of 2024 to January of
19 2025. He has just said they did not produce any insight
20 reports from post May of 2024. So it can't be that they have
21 given us anything that they are saying that they've given us.

22 MR. MUTTALIB: If I could just correct one statement.
23 We never claimed that these reports reflect the responses to
24 those surveys. It's about the underlying issue. Audience
25 reports are about the underlying issue of traffic and decreased

P8CJOPE3

1 traffic to the Times's website. AIG reports are about the
2 effect of generative AI on the Times. And it's the Times's own
3 analysis we think that is better evidence of the Times's own
4 understanding of its business. That is the Times's analysis
5 produced in its own internal reports, rather than the Times
6 speculating on survey responses by other people and trying to
7 determine what those survey responses mean. We think the
8 survey speaks for itself.

9 THE COURT: Back to the AIG reports, the date range
10 that has been produced was July 2023 to when?

11 MR. MUTTALIB: The same end date, your Honor.

12 THE COURT: May 2024?

13 MR. MUTTALIB: That's right.

14 THE COURT: When was the Times's lawsuit filed?

15 MR. MUTTALIB: December of 2023.

16 THE COURT: Okay. All right. So we've got two
17 categories of documents that you say go to showing the
18 understanding of the New York Times's own business and the
19 effect of generative AI, and those are audience reports that
20 have been produced, decreases in traffic that have been
21 produced from sometime in 2022 through May of 2024, and then
22 also AIG reports which speak directly to impacts of GenAI on
23 New York Times, and those are from July 2023 to May 2024.

24 Where is the disconnect, or are you still going
25 through the buckets of documents you've produced?

P8CJOPE3

1 MR. MUTTALIB: Your Honor, I wanted to save the Court
2 time, but I'll go ahead and list four more categories of
3 documents. First, we agreed to produce and have produced
4 documents regarding reductions in readership, revenue, or
5 online subscriptions. These are custodial documents.

6 THE COURT: I'm going to call those custodial
7 documents regarding --

8 MR. MUTTALIB: Reductions in readership, revenue, and
9 online subscriptions.

10 THE COURT: Okay.

11 MR. MUTTALIB: Second, custodial documents sufficient
12 to show the number of Times's subscribers from 2015 to the
13 present.

14 THE COURT: Okay.

15 MR. MUTTALIB: Third, custodial documents sufficient
16 to show the Times's monthly revenue, including from
17 subscriptions from 2015 to present.

18 THE COURT: Okay.

19 MR. MUTTALIB: And fourth, custodial documents
20 regarding decreases in online traffic or impacts on *Wirecutter*
21 and *Cooking*, so the Times's affiliate sites, and subscriptions
22 and advertising revenue.

23 THE COURT: All right. So what's the survey -- I
24 notice the word "survey" doesn't show up in your bucket. So I
25 understand there was one survey done, a survey of 15,000

P8CJOPE3

1 readers, okay. And when was that survey done?

2 MR. MUTTALIB: That survey I believe covered 2022 to
3 2023. it was produced a while back. I don't actually know the
4 exact date.

5 MS. NAJEMY: Your Honor, the survey covered a time
6 period of June 2024 to January 2025 all of the documents that
7 Mr. Muttalib has just told you about end with a production date
8 of -- they end in May 2024. So there was no document that the
9 Times has produced that has any sort of report or analyses or
10 study of this particular survey that they ran. And this is the
11 only survey that the Court has told us is relevant as it --

12 THE COURT: Slow down. Survey, when was it done? I'm
13 hearing completely different answers from you.

14 MS. NAJEMY: They produced a spreadsheet of data. The
15 survey, as I understand it from that spreadsheet of data, ran
16 from June 2024 -- July 2024 until January 2, 2025.

17 THE COURT: Okay. Pause. So the information in that
18 survey, the information gathered from that survey of 15,000
19 subscribers --

20 MR. MUTTALIB: 15,000 canceled subscriber; that's
21 right.

22 THE COURT: 15,000 subscribers. Can we just use that
23 term? Is that accurate enough?

24 MS. NAJEMY: I would say former subscribers, as they
25 had canceled their subscription to the Times, but yes, your

P8CJOPE3

1 Honor.

2 THE COURT: Okay. 15,000 former subscribers, why did
3 you cancel? And the results of that survey – I feel like
4 Richard Dawson for those of you old enough to remember – is
5 that there's a miniscule number of former subscribers who said
6 GenAI was the exclusive reason for canceling?

7 MS. NAJEMY: Correct, your Honor.

8 THE COURT: All right. So all of the categories that
9 Mr. Muttalib has put out say stop in May 2024. And then
10 there's a survey of former subscribers that cover a time period
11 from July 2024 to January 2025 that says why did you cancel,
12 which you have. You have the numbers. And you want the
13 New York Times's own views or analyses about the results of
14 that survey.

15 MS. NAJEMY: Correct, your Honor. As we understand
16 the Court's prior rulings, the only cancellation survey that is
17 relevant to this case is this single survey because it included
18 "AI makes my subscription unnecessary."

19 THE COURT: I wouldn't necessarily say that I had
20 ruled that that's the only data that's relevant to this
21 request.

22 MS. NAJEMY: With respect to survey evidence, your
23 Honor, that has been produced by or that the Times itself has
24 gathered. Our understanding is that the other cancellation
25 surveys that did not relate to generative AI or AI are outside

P8CJOPE3

1 the bounds of discovery, at least for the moment. And so we
2 are focused on this single survey that the Times has run and
3 provided to us and understanding how the Times itself --

4 THE COURT: And you've gotten the data?

5 MS. NAJEMY: We do have the data, your Honor, but we
6 have no reports. So in its opposition the Times says we gave
7 them all these reports that explain our views about these
8 surveys.

9 THE COURT: I understand now with the timeframe. You
10 want reports related to that survey?

11 MS. NAJEMY: Correct, your Honor.

12 THE COURT: And I go back to my original question,
13 which is now that you have that survey, the results of that
14 survey, still again, why is this not an expert issue?

15 MS. NAJEMY: Again, your Honor, the Times's own
16 analysis of these survey results and what it has to say about
17 them and what it is saying about the impact of GenAI and
18 whether or not it is harming it and what it reads from this
19 survey, as it is in the best position to understand how this is
20 impacting its business and how it interprets these results,
21 these reports or analyses, they are the best evidence of that,
22 answering that question.

23 THE COURT: You're still not answering my question of
24 why is this not an expert issue? If there are analyses from
25 this, would the experts not -- I'm not going to put words in

P8CJOPE3

1 Mr. Muttalib's mouth, but talk to me Mr. Muttalib.

2 MR. MUTTALIB: Two responses, your Honor. First, as
3 I've already articulated, the survey has dozens of fields of
4 information, which takes me to the second point. They're going
5 to have their own expert who can analyze all of these fields
6 including the apparent lack of number of individuals who picked
7 generative AI subscriptions made -- generative AI made my
8 subscription unnecessary with the relevant demographic data of
9 those subscriber -- I've already listed those for you,
10 geography, gender, etc., and for all 15,500, their initial
11 reasons for subscribing and their reasons for unsubscribing.

12 Their expert can look at that 15,000-plus number of
13 data, times however many fields, dozens of fields, and come up
14 with their own conclusions. They will also have our own expert
15 report which will analyze the ways in which the Times was
16 harmed by defendants' products. That has to be sufficient.

17 MS. NAJEMY: Your Honor, if I may briefly respond.
18 This may very well be an expert issue, but the Times's own
19 admissions about how this is or is not impacting its business
20 are a relevant input to that expert analysis. So yes, they can
21 take a look at the survey and they can draw all the conclusions
22 they want, but what is also important and essential to that
23 analysis is what the Times has to say and how they are
24 interpreting it. And their admissions that in fact generative
25 AI is not impacting or harming their business or causing them

P8CJOPE3

1 to lose subscribers, that is an essential piece of the expert
2 analysis in this case.

3 THE COURT: Well, but it's an expert analysis.

4 MS. NAJEMY: This is an input to the expert analysis.
5 The Times's admission that this is in fact not harming their
6 subscriptions --

7 THE COURT: You have that information in the survey.

8 MS. NAJEMY: No, we have the survey results which tell
9 us what subscribers have told the Times. But how the Times is
10 reading those results and what they have to say about it,
11 whether they believe it was defendants' generative AI products
12 versus a different set of generative AI products, all of those
13 things are relevant to the harm that the times is claiming here
14 and blaming on the defendants in this case. What these
15 analyses say, those admissions from the Times are incredibly
16 relevant to the fair use analysis here.

17 MR. MUTTALIB: Your Honor, those are great questions
18 for a deposition.

19 THE COURT: So here's my ruling. Premature. Ask at
20 deposition, and you can reraise it during expert discovery if
21 you want, but nothing is stopping you from going out and doing
22 your own surveys if you want.

23 MS. NAJEMY: Thank you, your Honor.

24 MS. BROOK: Your Honor, the next one, the sort of the
25 bridge, as I mentioned, between news and class is actually on

P8CJOPE3

1 Exhibit C of the chart. It's Docket Number 390. New York
2 Times is joined by some of the other plaintiffs' motion
3 regarding open AI's text and social -- deficient text and
4 social media production.

5 THE COURT: 390, okay. What I understood was that
6 there was some meeting and conferring with OpenAI on August 4.
7 So maybe I should ask OpenAI to address this first, or is this
8 issue resolved.

9 MS. BROOK: We were at the meet and confer, your
10 Honor, so I'm happy to -- I was not personally, I will say, but
11 New York Times was and I got the full report, and the issue is
12 not resolved. And I am so reluctant to go back to meet and
13 confers without clear guidance from this Court because, your
14 Honor, let me say why. This is the issue that is going to
15 prejudice the plaintiffs for the next six months of discovery
16 if it's not resolved promptly.

17 We've met and conferred a couple of Times, and there
18 is a couple core fundamental disagreements about what OpenAI is
19 required to do when it comes to their production of text and
20 social media messages. And for the record, when I say "text,"
21 I'm referring broadly to any app-based messaging platform. And
22 the disagreements are this, your Honor, and there are
23 disagreements that I thought my counsel for the class
24 plaintiffs and counsel for the New York Times had moved on in
25 October and November and this Court ruled on both at the

P8CJOPE3

1 hearing and in December and that were resolved precisely so
2 that we wouldn't find ourselves on the eve of depositions
3 without full text and social media message.

4 These are the three issues, your Honor. Conduct
5 proper custodial interviews of their current employees. I
6 would love a straight answer today from counsel for OpenAI
7 because we're not getting it on the meet and confer. Who
8 conducted the custodial interviews? Because it of course
9 should be counsel of record in this action and not personal
10 counsel. We have no objection to the existence of personal
11 counsel, but it is not they that should be conducting the
12 custodial interviews.

13 And what question was asked to ensure that they got
14 correct answers from lay witnesses about whether or not they
15 had potentially responsive information? I will be very clear.
16 My understanding of what the Court ruled in December is that
17 they should be asked whether or not they had messages related
18 to this litigation, and whether or not it was on a work or a
19 personal phone was not relevant to whether or not it should be
20 collected and produced.

21 We, the New York Times, asked our people three
22 questions because, again, lay witnesses are not those that have
23 the knowledge to make these decisions. So we said did you use
24 your phone for work? Did you text for work? Did you use
25 WhatsApp appear for work? Did you use Signal for work? Did

P8CJOPE3

1 you use all of those same things? And it's a much longer list
2 of specific things that we asked our people to communicate
3 about this litigation.

4 And then same thing, did you use any of those specific
5 applications or social media to talk about generative AI? Now,
6 I'm not asking OpenAI to do all three of those things because I
7 think asking them whether they use text messaging to talk about
8 generative AI is probably a little bit too far. But very
9 simple what we want, your Honor, that they either represent to
10 the Court that they already did this or that they do it right
11 away, that counsel for record interview current custodians, ask
12 whether or not they used a variety of text messaging apps or
13 social medias to communicate about their work or to communicate
14 about this litigation.

15 If the answer to that question is yes or maybe, then
16 those applications need to be collected and searched with the
17 agreed-upon search terms in this case. So that's step one,
18 just conduct a proper custodial interview. And if they've
19 already done this, this may be quick. But despite asking these
20 very direct questions, we have not gotten direct answers. Then
21 we need to do a proper custodial --

22 THE COURT: No, we need to get the three issues out,
23 and then I'll let you respond to the three issues and put
24 additional issues of disagreement if you want to. I get that
25 this is an extremely sensitive issue, but I really want to get

P8CJOPE3

the issues out before we get distracted. Go ahead.

(Continued on next page)

P8CEOPE4

1 MS. BROOK: And I would be delighted to be corrected
2 if I am wrong about any of this, your Honor.

3 Issue two: Conduct a proper custodial collection and
4 review. Again, it's simple. If there's any doubt as to
5 whether a given custodian used their text or social media to
6 communicate about issues responsive to The Times' request, then
7 the devices should be collected, the agreed-upon search terms
8 should be run, and responsive documents should be produced or
9 logged. And again, we had assumed, but perhaps we need to say
10 explicitly, all of that work should be done by counsel of
11 record in this action.

12 Most significantly, your Honor, and this is what we
13 learned --

14 THE COURT: When you say it should be done by counsel
15 of record, that means not personal counsel of particular
16 individuals or executives.

17 MS. BROOK: That's the exact distinction I am making,
18 your Honor.

19 THE COURT: Okay.

20 MS. BROOK: Most critically there, we learned that
21 personal counsel did an initial review of Mr. Altman's and
22 Mr. Brockman's text messages before it went to counsel for
23 OpenAI in the first place.

24 Now, what counsel for OpenAI is going to tell you is
25 they were instructed -- personal counsel was only instructed to

P8CEOPE4

1 remove highly sensitive, personal-type messages. No. Wrong
2 order. OpenAI should do the responsiveness review first and
3 foremost. Then if they want to give the subset of documents
4 that they intend to produce to personal counsel to do a privacy
5 review, that's fine.

6 But that gets me to the third bucket, your Honor,
7 which is appropriate privacy redactions, and I want to make
8 sure we are all on the same page here. Something can be highly
9 personal, highly, highly personal but responsive. That gets
10 produced. There can be two individuals at a company, as we all
11 learned from Coldplay that are having an affair, and that's
12 highly personal and private, but if they are talking about
13 issues relevant to the case where it's also clear that they're
14 involved in a relationship, that's responsive and needs to be
15 produced. That's what I mean.

16 So, we need the order to be review by counsel of
17 record at OpenAI for responsiveness, and then if they want to
18 do a privacy review with personal counsel, that's fine, they
19 can take out things that are not responsive in any way, and
20 fall into an agreed-upon bucket which we don't yet have, but
21 that I am confident we should be able to meet and confer about,
22 of appropriate privacy categories like information about minor
23 children or health issues or things like that.

24 Those are the three main issues, and I'll just
25 summarize them again: Conduct proper custodial interviews,

P8CEOPE4

1 conduct proper custodial collections and review, and make any
2 appropriate privacy redactions.

3 The last issue I'll flag is much more minor, but
4 again, even after multiple meets and confers we couldn't get
5 agreement on this. Maybe the Court will have a different view
6 on it than I do, but I kind of found their position surprising.
7 When they first produced the 35 messages that they produced
8 after eight months of review, which was their total production,
9 they redacted phone numbers of the nonperson. So, Mr. Altman
10 is talking with someone and so they redact the other phone
11 numbers. Now, they've agreed to unredact the phone numbers.
12 Great. Of course you have to unredact the phone numbers. Of
13 course. But we also said, you have to unredact them going
14 forward and any other productions. And if it just shows up as
15 a phone number, right, Mr. Altman is texting with someone at
16 650-blah-blah-blah-blah-blah-blah-blah, then you should ask
17 Mr. Altman to look up in his phone who this person is so that
18 we all know and just send us an email. Right? This phone
19 number corresponds with X, Y engineer at OpenAI, or this phone
20 number corresponds with Mr. Altman's college roommate.

21 They're not agreeing. And again, if they want to
22 correct me or change their position, I welcome it. They are
23 not agreeing to provide that information going forward.
24 They've agreed to provide it for the 35 lovely text messages
25 they've produced so far but not in the future, and this is not

P8CEOPE4

1 something I think we should be fighting about over and over
2 again.

3 So, that's it, your Honor: Proper custodial
4 interviews where they're asking, "Did you use these specific
5 applications for your work, did you use them for issues related
6 to this litigation?" Because again, custodians should not be
7 the ones entrusted to know what's responsive to the RFPs that
8 The New York Times served, just like The New York Times
9 custodians aren't the ones making calls about OpenAI's RFPs.
10 Collection search terms apply. Review, in the first instance,
11 by counsel for OpenAI. It doesn't need to go in-house at
12 OpenAI if there are privacy concerns, they can treat it as AEO.
13 And then agreed-upon privacy redactions that do not exclude any
14 responsive materials just because it may also be highly
15 personal.

16 And then with my little sub note of let's all tell
17 each other what phone numbers we're talking about, because if I
18 get to a deposition and I ask Mr. Altman who is
19 650-what-have-you, he's not going to know, and that's not his
20 fault.

21 THE COURT: Mr. Gratz, come on up.

22 So, I have proper custodial interviews of current
23 employees, proper custodial collection and review, and
24 appropriate redactions, and then the fourth and minor issue is
25 this issue of phone numbers and agree to say who they are going

P8CEOPE4

1 forward.

2 All right. Mr. Gratz.

3 MR. GRATZ: The first thing I would like to do is turn
4 to ECF 390, the letter brief that doesn't raise any of these
5 issues. We've got a number of new things that the parties have
6 been meeting and conferring about. I would be delighted to
7 give the Court an update on that meet and confer. We have a
8 further meet and confer scheduled for -- it was going to be
9 tomorrow but counsel for plaintiffs wasn't available, so we're
10 doing it Thursday or Friday, but --

11 THE COURT: You're all spending the week in New York
12 basically?

13 MR. GRATZ: Well, we may be doing it remotely.

14 These are issues that the parties are actively meeting
15 and conferring about. I would be delighted to meet and confer
16 further with plaintiffs. I can provide an update on that meet
17 and confer on the record. But these issues that Ms. Brook is
18 raising are not in the papers, and we have not had an
19 opportunity to brief them. The parties should go back and talk
20 about them, because I think Ms. Brook will find there is less
21 disagreement than Ms. Brook thinks about these issues. Not all
22 of them, but the disagreements are at a level of granularity
23 that I don't think the parties have joined issue on in the
24 papers thus far.

25 THE COURT: Okay.

P8CEOPE4

1 MR. GRATZ: I would be happy, again, to address any of
2 these issues that the Court is interested in, but there is
3 nothing for the Court to do today because the plaintiffs have
4 not asked the Court to do it. If the Court has questions that
5 the Court is interested in, I'm happy to answer them.

6 THE COURT: You know what, we're all here, we're going
7 to talk about some of this.

8 What I read from the summary of ECF 390 on the chart
9 -- which, by the way, the hyperlinks don't work. They go to
10 different documents, which is not helpful. So Exhibit C is
11 where the link to 390 is; right?

12 MR. GRATZ: That's right, your Honor.

13 THE COURT: I'll go to the docket myself and look it
14 up instead.

15 MR. GRATZ: I would be happy to hand up a copy if your
16 Honor would like.

17 THE COURT: Nope, I've got it. Or should I look at
18 the filing from the 4th?

19 MR. GRATZ: So --

20 MS. BROOK: I have a suggestion.

21 I would look at docket number 309, which is the motion
22 we brought in December of last year, which was argued, which
23 was ruled on. Sorry, November of last year. The order was in
24 December of last year. And I will read from that:

25 Apply the agreed-upon search terms and any subsequent

P8CEOPE4

1 terms to any collected messages. Text messages should be
2 produced in 24-hour increments. The parties can redact
3 portions of messages only for nonresponsive and highly
4 sensitive information, such as medical information.

5 This docket number, 309, from November 18, 2024, along
6 with the class plaintiffs' motion, which is docket number 230,
7 from October 23, 2024, raised pretty much, if not every issue I
8 just went through, basically every issue. And it was argued
9 and it was ruled upon. And OpenAI then took 300-plus days to
10 produce a single text message. It did produce 35 text messages
11 from three custodians and zero social media messages. So I
12 take issue with the argument that we haven't joined. We
13 joined, we argued, it was resolved, and now I'm back. And I --

14 THE COURT: It's raised again. I mean, I see some of
15 it in 390 also, no?

16 MR. GRATZ: Your Honor, that's right. I agree that
17 that is what the Court ruled, and we have complied and are in
18 the process of continuing to comply with that order. As the
19 papers on this motion identify, there's some issues with
20 respect to collection and production for two custodians that
21 the plaintiffs have brought to our attention that we are in the
22 middle of rectifying. We have made additional productions in
23 the last few days, we are continuing to do so. And it is about
24 that issue that the parties have this meet and confer later
25 this week to close out those remaining issues.

P8CEOPE4

1 I'm happy to address these issues one by one.

2 With respect to the redaction of phone numbers, I
3 think the parties have reached agreement. We are providing the
4 names of the phone numbers when the person is not otherwise
5 identified. And if there's just a phone number, that person
6 isn't in the phone. And if they want to know who it is --

7 THE COURT: That's not necessarily the case, and I can
8 attest to that having just gotten a new phone.

9 So, next.

10 MR. GRATZ: Well -- all right, your Honor.

11 THE COURT: That's an inference one may draw, but I
12 would say that is not the only conclusion one could draw from
13 that.

14 MR. GRATZ: I am not intending here to draw an
15 inference, your Honor, but to report our understanding of how
16 the technical collection process works. That is, where that
17 person is in the address book, there's a name in what we
18 collect and produce. And where there isn't, there isn't. I
19 mean, if there are particular phone numbers that plaintiffs
20 want to meet and confer about, we're happy to, but --

21 THE COURT: No, no, no, no, no, no, no. What I heard
22 earlier was if there's only a phone number, you'll provide the
23 names associated with the phone number.

24 MR. GRATZ: That's not correct, your Honor. We may
25 not know the name associated with that phone number.

P8CEOPE4

1 THE COURT: Well, then you find out; right?

2 MR. GRATZ: Well, they can find out through a
3 discovery mechanism that is not --

4 THE COURT: No, no, no, no, no, no.

5 MR. GRATZ: -- request for production --

6 THE COURT: You find out. No. You can find out.
7 We've been dancing around this long enough.

8 MR. GRATZ: Understood, your Honor. So we're happy to
9 do an investigation to find out for any phone number that the
10 plaintiffs would like us to.

11 MS. BROOK: Every phone number in a text message
12 produced by the defendants in this action.

13 THE COURT: If you want me to rule on a chart and
14 direct you to provide that information under seal to the Court,
15 I will.

16 But let's get back to the -- that was supposed to be
17 the minor issue.

18 MR. GRATZ: Yes. And --

19 THE COURT: What about the other three?

20 MR. GRATZ: Right. So, I think we are largely aligned
21 on the major issues. That is, with respect to the collection
22 review issues, we have done appropriate custodial interviews.
23 What was said in those and how those were done is an issue that
24 is privileged, but I'm happy to share that they were done by
25 company counsel. And I think they got at exactly what

P8CEOPE4

1 Ms. Brook's questions got in their entirety.

2 And I think that simply closes out that issue.

3 MS. BROOK: So I'm pleased to hear they were done by
4 company counsel. That is not an answer we've gotten yet and
5 it's welcome news.

6 On the privilege claim, I mean, just come on. Let's
7 all work together here. Can they say whether or not they asked
8 their lay people, "Did you use your phones for work?"

9 MR. GRATZ: Yes. We said that.

10 MS. BROOK: Okay. And if they said they used their
11 phones for work, "All right, we might have used our phones for
12 work," were they collected?

13 MR. GRATZ: My understanding is that the answer is
14 yes.

15 MS. BROOK: If that's their representation, then
16 that's great.

17 MR. GRATZ: There's no reason this -- well --

18 MS. BROOK: That's what we want. Again, I'm happy to
19 get the answers. Delighted, in fact. So, if OpenAI is saying
20 to the Court that they conducted all of their custodial
21 interviews by counsel of record in this case and that at those
22 interviews --

23 THE COURT: Well, I heard "company counsel," which may
24 not be the -- is that --

25 MS. BROOK: Oh, thank you. That's a great point.

P8CEOPE4

1 Surely, it had to be done not by in-house counsel alone.

2 MR. GRATZ: Your Honor, the level of detail on what
3 lawyers working for OpenAI conducted those particular
4 interviews is information that I don't have today.

5 THE COURT: Okay. So, why don't we do this: We can
6 get a declaration by counsel after your meet and confer. Let's
7 get it on Friday. This one is not going to be subject to the
8 Ninth Circuit argument. I'm sorry. But a declaration
9 answering those questions. It should all be in the transcript.
10 If you want to file it under seal, you can.

11 MR. CONNERS: Your Honor, Jordan Connors for class
12 plaintiffs.

13 I was on the meet and confer so I want to clarify one
14 point.

15 THE COURT: Okay.

16 MR. CONNERS: On the were these documents collected,
17 there is a difference between having them collected by personal
18 counsel and screened for whether they are business or personal
19 and then passing them on to counsel of record, or if somebody
20 uses a device for business, producing all of the documents that
21 hit the agreed search terms or at least producing them to
22 counsel of record for responsiveness review. And I just want
23 to make sure we don't miss that distinction, because as I
24 understand what OpenAI represented at the meet and confer was
25 those documents were first screened by personal counsel before

P8CEOPE4

1 they were provided to counsel of record.

2 MR. GRATZ: I'm happy to address that, your Honor.

3 THE COURT: Thank you. Please do.

4 MR. GRATZ: There is nothing inappropriate about the
5 ordering of having work messages sent to company counsel.
6 Anything that has to do with work, not things that do not have
7 to do with work, and for company counsel to make a
8 responsiveness determination on their own and make a
9 production. It is not, by any means, required that an
10 individual turn over the entire contents of their phone to
11 their employer, period. It is appropriate for the work stuff
12 to go to the employer, and it is not appropriate for the
13 not-work stuff to go to the employer.

14 MS. BROOK: So now we're right back where we were in
15 the fall of 2024. We had this fight. I can guarantee you,
16 your Honor, that Mr. Gratz and I disagree about what's a work
17 message and what's not a work message. I am not a Silicon
18 Valley tycoon myself, but I would posit that Mr. Altman texting
19 with other Silicon Valley tycoons about GenAI and its impact on
20 the space and all of this is deeply relevant to his work and
21 should be produced in this case, and there's no basis for
22 drawing some magical distinction between the work and the
23 personal.

24 So, the simplest way to do this, which is, again, what
25 I thought we had resolved very long ago, is this, and I'm happy

P8CEOPE4

1 to prepare the questions that we want specific answers to by
2 the end of the day tomorrow and send it to OpenAI so they can
3 provide their sworn answers to the Court on Friday. But the
4 simple process should be this, and I appreciate Mr. Connors
5 pointing out yet the other wrinkle:

6 Outside counsel of record should attend and, frankly,
7 manage the custodial interview. If in-house counsel wants to
8 attend as well, God bless, but it should be conducted by
9 outside counsel of record, and there should be clear questions
10 as to whether the lay witnesses use any sort of text messaging
11 or social media application to communicate about their work or
12 this litigation. If the answer is yes or perhaps, it should be
13 collected by outside counsel of record. Again, if they want to
14 treat it as AEO internally, you know, even with their own
15 clients is what I'm saying, fine. But there's no reason why
16 Mr. Altman should not trust Mr. Gratz here to then apply the
17 search terms, the agreed-upon search terms, to the documents.

18 At that point, if it gets sent to now Gibson Dunn,
19 because we now have another major law firm here, though of
20 course not subject to the Court's jurisdiction, that wants to
21 review those messages and propose redactions or even propose,
22 hey, we don't think this text message is really businessy, they
23 can log that too. Right? But that has to be the order. And I
24 will say right now that I don't think we should be drawing
25 these distinctions about business or not because this is all

P8CEOPE4

1 related to OpenAI's business, and we know that. If it's highly
2 personal, that will capture the nonbusiness; right? That's
3 what will capture it.

4 If Mr. Altman is having a text message about what to
5 have for dinner that night and he happens to say, hey, I
6 trained as a professional cook, and so it hit on our search
7 term "training," we're not going to get that. Fine. Okay?
8 But that needs to be the process. And it needs have happened
9 six months ago. And if it doesn't happen with true expediency,
10 then we are not going to be able to start to take these
11 depositions because we won't have the basic custodial documents
12 that we're entitled to, to do so.

13 MR. GRATZ: And to be clear, your Honor, we expect
14 they have already received many additional text messages as
15 part of this meet and confer process. We expect those text
16 messages to be fully substantially produced by this week or
17 next. This is an issue that we are addressing with appropriate
18 expediency at this point.

19 MS. BROOK: The issue they are addressing --

20 MR. GRATZ: I --

21 MS. BROOK: Oh, sorry. Go ahead.

22 MR. GRATZ: I'm sorry.

23 THE COURT: Don't talk over each other, please.

24 MR. GRATZ: With respect to the idea that the only
25 appropriate way to do this is for a person's entire device to

P8CEOPE4

1 be turned over to their employer, that is not the only
2 appropriate way to do this, and that is, in this circumstance,
3 not an appropriate way to do this under these circumstances.

4 I certainly agree that there are many interactions and
5 messages that are with people who are not OpenAI employees, not
6 people at work, that may be relevant. And we are producing
7 those, because those are not the highly personal material that
8 we are talking about here. But because these things are
9 getting produced in 24-hour chunks in part, where there is a
10 mix in that 24-hour chunk and there is only a portion that
11 relates to work, where someone talking to, you know, their
12 spouse or a friend and part of it relates to work and the rest
13 of it does not, it isn't appropriate for us to produce the part
14 that doesn't relate to work.

15 THE COURT: Well, that's what the redactions are for,
16 no?

17 MR. GRATZ: That's right. And it is appropriate for
18 those redactions to get applied by personal counsel.

19 THE COURT: I don't think that's --

20 MS. BROOK: It's appropriate for those redactions,
21 perhaps, to be comprised in part by personal counsel but not in
22 the first instance, your Honor. Not before OpenAI has done its
23 own responsiveness review.

24 And Mr. Gratz keeps using the word "relevance." I
25 trust he means "responsive," not "relevance."

P8CEOPE4

1 MR. GRATZ: I do.

2 MS. BROOK: But it should be, first, OpenAI's counsel
3 that reviews, not personal counsel.

4 And frankly, to the extent personal counsel is
5 redacting for highly sensitive personal but not responsive, it
6 should be logged. To the extent they're redacting for what
7 they believe is something that's -- I guess I don't even
8 understand how it would get here, which is why this makes no
9 sense and why they're just complicating this process. If the
10 text message chain is 24 hours long and includes a discussion
11 that is responsive and a discussion that isn't, Mr. Gratz's
12 team is going to redact the second part in and of themselves.
13 They're going to do that. It's already done.

14 MR. GRATZ: An --

15 MS. BROOK: They're just overcomplicating it. And
16 worse. And we're not just trying to be difficult here.
17 Personal counsel is far from the best equipped and, probably
18 better said, not equipped to be the one to be making calls
19 about what's responsive and what's not. They're not here.

20 MR. GRATZ: That's why they aren't doing it.

21 THE COURT: We are going to stop this argument right
22 now. This is my order: Whatever meet and confer you had
23 scheduled for tomorrow or the next day, let it go forward. I
24 had a representation that there are more text messages to be
25 fully or substantially produced by this week or next.

P8CEOPE4

1 MR. GRATZ: Yes.

2 THE COURT: Your deadline is August 22.

3 MR. GRATZ: Thank you.

4 THE COURT: Okay. So produce what you're going to
5 produce August 22.

6 I go back to my order that I was in the process of
7 directing before we got interrupted. By August 15, there is
8 going to be a sworn statement by counsel, and it can be
9 attorneys' eyes only, and filed under seal. I have to be able
10 to see it, though.

11 MR. GRATZ: Yes, your Honor.

12 THE COURT: You'd be surprised. I've had other cases
13 where parties filed something under seal and even the Court
14 couldn't see it. So, some sort of super sealing.

15 So, can be designated outside counsel eyes only, and
16 in fact I would suggest that we do that because this is so
17 sensitive, file it under seal on the docket, about the process.
18 Answering the questions.

19 I understand that there are two particular custodians
20 whose names have been mentioned. Are there others or are those
21 the two main ones?

22 MS. BROOK: Oh. It's applicable to every agreed-upon
23 custodian. I believe there are 32 agreed-upon OpenAI
24 custodians. The reasons we keep mentioning those two names in
25 particular are:

P8CEOPE4

1 One, they only produced for three custodians, so we've
2 got a subset to deal with.

3 And two, those were the ones where, just based on our
4 own 24-hour investigation, we were able to point out that
5 Microsoft had produced a whole bunch of text messages with
6 those custodians that OpenAI hadn't produced. But no, it
7 should apply to all 32 custodians. They should make the
8 representation that this is how they did the custodial
9 interview, this is how they did the collection, and this is how
10 they did the review.

11 THE COURT: But there's three custodians that have
12 risen to the top; right?

13 MR. GRATZ: Two.

14 MS. BROOK: That's only because those are the only
15 three custodians they produced from those.

16 THE COURT: Who are the custodians that text messages
17 and social media messages had been produced from?

18 MS. BROOK: Mr. Altman, Gregory Brockman, and the
19 third I have in the brief.

20 THE COURT: Altman, Brockman, and -- I'm sure nobody
21 wants their name repeated multiple times in this conference.

22 (Counsel confer)

23 MS. BROOK: Maroon Shetty.

24 THE COURT: Shetty?

25 MS. BROOK: Yes, I believe so.

P8CEOPE4

1 Yes. OpenAI's production contained text messages from
2 only three: Sam Altman, Greg Brockman, and Varun Shetty,
3 S-H-E-T-T-Y, and zero social media messages.

4 THE COURT: All right. So we're going to --

5 MR. GRATZ: Your Honor, I really hesitate to add a
6 wrinkle before your Honor is about to rule, but it is important
7 to, I think, what your Honor is about to rule, which is that
8 among the custodians, the 32 custodians, many of them are
9 former employees who we don't control.

10 THE COURT: Well, I was going to only get you to the
11 three custodians for which documents have been produced first.

12 MR. GRATZ: I --

13 THE COURT: No, no, no, no, no. I'm not saying you
14 don't have to do it for the rest, but these rise to the top in
15 order of importance and urgency; right?

16 MR. GRATZ: Sure.

17 THE COURT: So, you're going to provide that sworn
18 statement for at least these three custodians by the 15th.

19 Ms. Brook, what were you going to say?

20 MS. BROOK: I think there's a missing fact here. The
21 production that OpenAI made to us on July 14, I believe, of
22 this year, they said that's it for text messages. So they're
23 supplementing based on us elevating these specific issues. But
24 the reason I wanted, and I'm fine to limit it to all current
25 OpenAI employees who are custodians, but the reason I need it

P8CEOPE4

1 for all of them is that was their substantial completion
2 deadline for -- because we kept on saying we're going to come
3 to the Court because you guys aren't doing this quick enough,
4 so they self-imposed a substantial completion deadline of
5 July 14, and that was all we got, 35 from three.

6 So the reason I need it for the other delta is because
7 they're saying, well, they don't have anything responsive, and
8 I'm saying, well, is that because you messed up the custodial
9 interview process? Is that because you messed up the
10 collection? Is that because you messed up the review? Or is
11 that because they don't anything responsive?

12 THE COURT: Okay. So I'm going to stick with
13 August 15 for these three custodians. Sworn statement that
14 addresses who did the custodial interview and when, who else
15 was present, and whether the particular questions were asked as
16 far as apps and methods of communications, and whether work or
17 personal phones. So that's the first bucket of information
18 that will be in the sworn declaration.

19 The second bucket is how is the custodial collection
20 and review done. Again, we're going to start with those three
21 custodians because there's been some amount of documents
22 produced. Since we know that at least Mr. Altman used X --

23 MS. BROOK: Yes.

24 THE COURT: -- there needs to be some explaining of
25 why there has been zero messages from X produced.

P8CEOPE4

1 MR. GRATZ: And, your Honor, that won't be true by the
2 time we are done.

3 THE COURT: Okay. Great.

4 MR. GRATZ: So we think that it may make sense to
5 answer that question afterwards.

6 THE COURT: No, no, no. You're going to answer that
7 on the 15th for these three custodians: Altman, Brockman,
8 Shetty. Okay?

9 And that the third group of answers is whether and how
10 redactions or exclusions were made. The questions should be
11 pretty evident from the transcript. I will be looking at the
12 transcript when I am reviewing the filing on August 15 to see
13 if it complies.

14 As to the other custodians, it is going to be limited
15 to custodians who are, as of today, OpenAI employees out of
16 those 32. And you're going to answer the same questions, but
17 you're going to have until August 22 to do that. Okay? Same
18 kind of filing.

19 MR. GRATZ: Understood, your Honor.

20 THE COURT: If you're able to work out the issues --
21 no, I want to see the declaration on August 22 anyway. All
22 right?

23 MR. GRATZ: Understood, your Honor.

24 THE COURT: All right.

25 Next issue.

P8CEOPE4

1 MS. BROOK: We have gotten to the class issues, and so
2 that is Chart B, Exhibit B, which I will try to say what page
3 it starts on in a moment. They have only four, so they're
4 winning.

5 MS. BROCK: This is docket 378, which is OpenAI's
6 motion to compel their responses.

7 Eleanor Brock of Keker, Van Nest & Peters for
8 defendant OpenAI.

9 I'll try to keep this short because I think it should
10 be pretty straightforward. I'll just tell you what we're
11 seeking, why, and why it's proper under Rule 36.

12 What we're seeking here are straightforward admissions
13 to six RFAs asking what evidence plaintiffs have for their
14 theory that ChatGPT outputs substitute for their works in the
15 book market. For example, RFA 39 asks them: Are you aware of
16 any specific ChatGPT outputs that make full copies of your
17 asserted works? The answer to that is a fact, it's a yes-or-no
18 answer.

19 Same for RFA number 38, which asks them to admit
20 whether or not they're aware of any specific ChatGPT outputs
21 that contain more than 100 words of verbatim text. Again,
22 that's a fact. We just need a yes-or-no answer.

23 Other plaintiffs in this litigation from the MDL case
24 have already answered this, and this was the subject of a prior
25 motion to compel in this MDL at docket 131. They were

P8CEOPE4

1 originally objecting on the basis of attorney work product. In
2 the process of briefing that, they promised to substantively
3 response to the RFAs, and the Court ordered them to do so but
4 they didn't, and they are now asserting new objections,
5 claiming that they seek expert work product and that they're
6 improper under Rule 36, so we're back here litigating the same
7 issue.

8 Why would we need this? The idea that ChatGPT outputs
9 can serve as substitutes for plaintiffs' work is one of their
10 key market harm theories in this litigation. In fact, they
11 don't just claim that substitute outputs exist, they claim that
12 they are so prevalent that they are flooding the market for
13 their books and disrupting the job market for authors. They've
14 made this claim over and over again in their complaint, in the
15 consolidated complaint, recently at the tech tutorial. All
16 we're asking them is to admit or deny whether they have
17 specific evidence of this claim, and that's what these are
18 getting at.

19 Why are these proper? Well, first of all, they
20 already agreed to answer them, so we're just asking them to
21 honor that promise. But under Rule 36, these RFAs are designed
22 to narrow the issues in dispute, and that's exactly what these
23 do. What we're trying to do here is eliminate the categories
24 of evidence that we need to respond to and confirm what
25 evidence is still at issue in this case, and both of those

P8CEOPE4

1 things will be helpful for ensuring that our depositions and
2 expert discovery is productive and efficient.

3 Whether they are presently aware of such outputs is
4 also independently relevant because they're claiming that these
5 outputs are so prevalent that they're flooding the market. So
6 if plaintiffs themselves haven't encountered these outputs in
7 the nearly two years that we've been litigating this case, it
8 is highly relevant and it very much undermines their claim of a
9 flooded market.

10 THE COURT: Okay. Let's hear from plaintiffs.

11 MS. SALINAS: Good afternoon, your Honor.

12 Alijandra Salinas on behalf of the class plaintiffs.

13 Just a little bit of background and then I'll respond
14 substantively to the points.

15 First, plaintiffs fully complied with our previous
16 representation that we would amend our responses. We did so,
17 and because of that, there are four less RFAs before you than
18 there were on the prior motion.

19 THE COURT: So there were ten before?

20 MS. SALINAS: Yes.

21 THE COURT: Okay.

22 MS. SALINAS: Now, I want to really get to, I think,
23 the heart of the question. In their brief, they say they are
24 entitled to understand if plaintiffs' actual evidence of
25 alleged substitution. We do not disagree with that notion.

P8CEOPE4

1 The only question before this Court a matter of timing and
2 process. And we think it's pretty clear under the case law,
3 and in fact, don't just take our word for it. OpenAI's own
4 arguments about when RFAs are proper, as your Honor may recall,
5 one of the first motions class plaintiffs brought before you
6 was a motion to compel RFA responses regarding what was and was
7 not in the training data. That was an instance where everybody
8 had the data and the only question was for you to confirm,
9 which you could see right in front of you, was whether or not
10 that text was, in fact, a book.

11 OpenAI argued to you Rule 36 is not a discovery tool
12 to discover new information, and that RFAs were improper
13 because they were seeking to learn information that they do not
14 know. That is exactly what these RFAs are intending to do;
15 they're trying to figure out whether or not plaintiffs have
16 certain information. That is information that they contend
17 they do not know. And when OpenAI made that argument to you
18 last fall, you agreed with them, and so there is no reason to
19 reach different conclusion here. They are trying to seek
20 information that they do not know. That is improper by
21 OpenAI's own argument, and it remains improper here.

22 The second point I will make is there's a big concern
23 here of expert work product. The case we cited to the Court,
24 which they don't dispute, *Circa Bridge Loans*, very clearly
25 holds that when information and facts are known by a consulting

P8CEOPE4

1 expert, they are not discoverable. They do not dispute this
2 case, they do not offer an alternative case that says if facts
3 are known by consulting expert, they are discoverable. Again,
4 that's the end of the question here.

5 What we are saying is these are facts that are being
6 investigated by our consulting experts, and it is simply
7 improper at this stage to disclose them. They are going to
8 have every opportunity to dissect and go question what our
9 experts disclose in their reports and that is a proper time to
10 have this discussion.

11 I will further note that the timing of this motion is
12 a bit concerning when they are actively trying to prevent the
13 plaintiffs from getting access to their output logs. If they
14 want to know whether or not there are, in fact, output
15 examples, regurgitation, all they have to do is search for our
16 clients' names in their books in their own output logs. It is
17 wholly improper for them to try to refrain from giving us that
18 discovery and try to get our plaintiffs almost in a gotcha
19 moment. "Oh, what can you give us before we give you all the
20 information." That's not proper. This is clearly expert
21 discovery. We have answered RFAs but we can do so outside of
22 expert discovery, but it is not proper to answer those RFAs
23 right now.

24 I'm happy to answer any other questions the Court may
25 have.

P8CEOPE4

1 THE COURT: Let me hear from OpenAI, and specifically
2 on the issue of if it this information is currently under
3 investigation by the consulting experts. Does that change the
4 response?

5 MS. BROCK: I'll address that first and then I'll
6 address the other arguments that she made about timing.

7 No, it doesn't change the analysis whether their
8 experts are separately investigating this issue. And I want to
9 be crystal clear about what we're seeking here. We are seeking
10 admissions about plaintiffs' own personal knowledge. As I
11 explained earlier when I was going through the relevance, what
12 we want to know is if plaintiffs themselves have encountered
13 these outputs in the market in the time that they've been
14 litigating this case. That's highly relevant to the strength
15 of their market harm theories.

16 What we're seeking here is facts which are not
17 work-product protected. We're not seeking expert analysis, in
18 fact we're not going after anything that their experts have.
19 All we want to know is if plaintiffs have this evidence,
20 whether they are separately investigating it with their
21 experts, and they're going to present that in expert discovery.
22 It is really not relevant and it shouldn't change the analysis.

23 And as to the timing argument, plaintiffs are not
24 making the argument that seems to me that the information
25 sought by these RFAs is relevant, they're basically just taking

P8CEOPE4

1 issue with the timing and with the idea that we're using this
2 as a discovery tool. And I'll just point out two things:

3 One, is that in their opposition, they cite to RFPs
4 that OpenAI served on plaintiffs as if they somehow render our
5 RFAs improper, but that's the whole point. We've asked for
6 this information. It doesn't appear that they've produced any
7 of these outputs. Our conclusion is that they don't exist, and
8 now we want them just to admit that fact.

9 And on their other supplements, the Court's order and
10 our original motion covered all of these RFAs. So the idea
11 that they complied when they just made new objections to half
12 the RFAs in the set is not correct. But it's helpful for the
13 Court here because if you look at their supplements -- and
14 that's Exhibit A to our current motion -- you'll see that where
15 I guess they now have evidence, they've gone ahead and denied
16 the RFA. It's only the ones where I guess their experts are
17 still investigating that they refuse to answer. But that makes
18 the exact point that I'm trying to explain, which is that these
19 RFAs are absolutely proper and well-timed and they're
20 answerable, and we are entitled to answers.

21 I'm happy to answer any questions.

22 THE COURT: I guess my question for plaintiffs is if
23 the RFA explicitly or is now taken to explicitly exclude
24 anything that has gone to or is known or is investigated with
25 the consulted experts, what's the problem with answering it? I

P8CEOPE4

1 mean, even if the answer is no, because they haven't done their
2 own investigation, they're not on the internet every day
3 looking for copies of their own works, what does it matter?
4 Why not?

5 MS. SALINAS: So, I would say two things:

6 One, throughout this case and in preparation of this
7 case, all of these investigations have been undertaken with
8 consulting experts and counsel. So there is not going to be a
9 universe of things outside of work that is protected by expert
10 work product.

11 The second thing I will say is OpenAI has absolutely
12 no response to the fact that they took a totally opposite view
13 on how RFAs should be handled when it applied to them. They
14 have no explanation because there is none other than they think
15 one set of rules should apply to them and a different set of
16 rules should apply to us. They do not dispute the fact that
17 they are using an RFA to understand new information. They can
18 ask this question in a deposition, that is a proper mechanism
19 to do so, but it cannot be that all of a sudden it's okay to
20 learn new facts when OpenAI wants to ask a question but not
21 when class plaintiffs want to ask a question.

22 THE COURT: One more response.

23 MS. BROCK: There is absolutely a universe of
24 information that's not covered by the expert consultant
25 privilege, and that's what's in plaintiffs' own heads as the

P8CEOPE4

1 ones who are makes these allegations at every step of this case
2 that ChatGPT outputs are flagged in the market. And so what I
3 think is proper to do, if they are still doing investigation
4 with their experts, is for them to give us a qualified
5 admission to the RFA. So we give this example in our motion.
6 Admitted except as to, you know, information that's solely in
7 the possession of expert consultants, and that leaves open the
8 possibility that they can then do that analysis and put forth
9 an expert report, but there's no justification for them to
10 withhold that information now.

11 And as to the idea that we're taking a contrary
12 position from our prior position on RFAs, I don't think that
13 that's true at all. The prior RFAs, which plaintiff propounded
14 were clearly fact finding -- you know, if you go back and look
15 at the motion, they're asking a series of RFA couplets. You
16 know, admit that this is in the training data, admit that it's
17 not in the training data. They're clearly seeking facts. Same
18 for all of the other cases that plaintiffs cite. These RFAs
19 are distinguishable because, first of all, there's only six of
20 them and there's not 900-plus.

21 THE COURT: Well, there's six of them but they pertain
22 to all of the named plaintiffs? All of the plaintiffs in the
23 class? I'm trying to understand.

24 MS. SALINAS: Correct.

25 And, your Honor, I might just say, the proposal she

P8CEOPE4

1 just offered, which is if we answer the RFA to acknowledge what
2 we had stated in the complaint, which is public recording about
3 flooding the market, and state that the rest is extra work
4 product, we're more than happy to provide that in an RFA
5 response.

6 MS. BROCK: If they're willing to admit all of these
7 RFAs that are subject to motion, then that's great.

8 THE COURT: Here's what I am going to do:
9 Plaintiffs, you can go provide your qualified
10 admission to use OpenAI's term. When can you do that by?

11 MS. SALINAS: So, just to confirm, your Honor. If the
12 admission is what I just described, we can provide that
13 information by the end of this week.

14 THE COURT: Okay. August 15.

15 MS. BROCK: It would be the qualified admission that
16 we're asking for in our motion, which is to say admitted to
17 each of these RFAs, we admit we are not aware of specific
18 ChatGPT outputs in --

19 THE COURT: You're not writing the response for them;
20 right?

21 MS. BROCK: Sure. I just wanted to clarify.

22 THE COURT: No.

23 MS. BROCK: Yeah.

24 THE COURT: You know what, their qualified admission
25 is going to be the qualified admission. And if nothing else,

P8CEOPE4

1 it narrows the issues. Okay?

2 MS. BROCK: Understood.

3 THE COURT: The other option, I was going to say, is
4 you can wait until after depositions.

5 MS. BROCK: Understood.

6 THE COURT: Okay? Because I do see there's some
7 overlap with what the consulting experts may know or may have
8 discussed or may have asked and discussed with particular
9 individual plaintiffs that may make the type of qualified
10 admission that you are trying to write for the plaintiffs
11 impossible for them to get. Okay? So that's up to them.

12 MS. BROCK: Understood.

13 THE COURT: All right. Next issue.

14 MS. BROCK: Thank you, your Honor.

15 MR. CONNERS: Good afternoon, your Honor.

16 Jordan Connors for the class plaintiffs.

17 I'm going to be addressing docket 388 on your chart.
18 It is another privilege issue. This one is differently
19 situated, though, so I want to --

20 THE COURT: There were clawed back; right?

21 MR. CONNERS: That's exactly right. There are really
22 two reasons these documents are a little bit different than
23 some of the other privilege issues that I know you've dealt
24 with.

25 So there are two categories: The first category is a

P8CEOPE4

1 set of originally ten documents which we had been seeking as a
2 sample for in-camera review to get a ruling from your Honor,
3 hopefully, that could guide the parties for a great number of
4 privilege disputes, and we have been seeking that since
5 February 19. We've had many meet and confers, this is the
6 third motion that we have brought. And it was ten documents as
7 a sample, and we chose this sample because it's from a universe
8 of documents where there is no lawyer on the communication. It
9 appears to be a business discussion, either because it's a
10 redaction and you can see from the rest of the context that
11 it's a business discussion, or something on the log identified
12 it as a business discussion, or there was just no
13 identification of any lawyer who gave privileged legal advice.
14 So it's a sample of ten documents.

15 We filed, for the third time, our motion, I believe it
16 was July 30. And then on August 4, OpenAI produced half the
17 documents that we had been fighting about, and these are
18 documents that had claimed their privilege log, established
19 that they were privileged, they submitted declarations. But
20 nevertheless, we filed our motion, and half of them come back
21 to us. So we, for the first time, can look at the underlying
22 documents that they have been describing on these privilege
23 logs, and of course OpenAI could not produce the documents
24 until we had filed our motion, so we couldn't put any of that
25 information in our brief. But suffice to say it confirms our

P8CEOPE4

1 suspicion that there are a great many documents that OpenAI has
2 been withholding that are actually business discussions among
3 business people, and even the business people might use the
4 word "legal" or might even have, in a business context, some
5 discussion of a legal issue, but there is no underlying
6 confidential legal advice from an attorney, and so they
7 shouldn't be withheld.

8 There is one case that I would love for your Honor to
9 focus on if you do indeed review these documents for privilege.
10 It is called *Reino de España v. The American Bureau of*
11 *Shipping*. I think that's Kingdom of Spain versus the American
12 Bureau of Shipping. It is a case in the context of a shipwreck
13 off the coast of Spain, and Spain sued the American Bureau of
14 Shipping because they didn't have enough safety standards, and
15 the American Bureau of Shipping inadvertently produced a bunch
16 of emails, including general counsel and a lot of business
17 people, that talked about what are the potential implications
18 of various safety standards, including the potential liability
19 of having various levels of safety standards. And the party
20 that received those documents submitted them to the Court for
21 *in-camera* review, and the Court found that even though there
22 was some legal discussion, it was all in the context of
23 competitive advantage or sort of what are the business
24 implications of various safety standards for this ship.

25 Very similar to documents we've seen produced that

P8CEOPE4

1 were previously held back on a claim of privilege where
2 business people are discussing potential legal implications of
3 various, you could call them, safety standards or compliance
4 standards with regard to copyright or other legal issues.

5 THE COURT: Okay. I'm going to pause you only because
6 we're already an hour past the scheduled end time for this
7 conference.

8 MR. CONNERS: Fair enough.

9 THE COURT: So, you said there were two categories.
10 That's the first category, is not down to five documents at
11 issue?

12 MR. CONNERS: That's exactly right.

13 THE COURT: Okay. What's the second category?

14 MR. CONNERS: Category two, I think, is much easier.
15 Category two are clawbacks.

16 THE COURT: Okay.

17 MR. CONNERS: It's narrow because the only documents
18 in this category, obviously, were produced by OpenAI in the
19 first instance, they were clawed back, typically on the eve of
20 a deposition or during a deposition. And under
21 Rule 26(b)(5)(B), plaintiffs have a right to bring them before
22 your Honor for a determination of whether they are privileged.
23 We have met and conferred on this topic many times now, and
24 consistent with those meet and confers and in OpenAI's brief
25 for your Honor, OpenAI has submitted zero authority, nothing,

P8CEOPE4

1 to suggest that the language of Rule 26(b)(5)(B) doesn't mean
2 exactly what it says, and I'll read it. It talks about
3 clawback procedures, then it says:

4 The party receiving a document clawed back for
5 privilege -- and now I'm quoting -- "may promptly present the
6 information to the Court under seal for a determination of the
7 claim."

8 And if you look at the advisory committee notes, I'm
9 not going to go into it because I know we're short on time, but
10 rest assured if you look at the advisory committee notes for
11 that exact rule, those are the 2006 amendments, that also makes
12 clear what the rule itself says, is that the party receiving
13 those documents has the option to submit them to the Court for
14 *in-camera* review. There is no authority going against us on
15 this point, OpenAI has certainly not provided anything, so we
16 think that should be open and shut.

17 THE COURT: Okay. Usually the clawbacks aren't an
18 issue, right, because usually they're clawed back and everybody
19 kind of goes, oh, yeah, I guess they were privileged, but here,
20 you're saying that they weren't.

21 MR. CONNERS: Exactly right. And we are certainly
22 not --

23 THE COURT: Okay. How many documents are you talking
24 about? Is this the 18 documents?

25 MR. CONNERS: This is the 18 documents in category

P8CEOPE4

two.

THE COURT: All right. Let me hear from defendants.

You know where I'm leaning and the numbers of the documents.

MR. GRATZ: I want to begin where my friend concluded, which is are they saying that these 18 documents aren't privileged? They are not saying that. They are making no argument whatsoever about that. And the bulk of these documents were produced, not withheld, but in redacted form. They are not making the argument to which I would love to respond, your Honor, that they are making as to the other documents, that either the log or the document itself shows that it's anything other than a privileged communication. They have not an iota of argument that those 18 documents are anything other than totally privileged. They have not articulated that argument at all. The argument that they have articulated for the other documents that I would like to address.

THE COURT: The five documents.

MR. GRATZ: The five documents.

For the 18 documents, they are saying they were clawbacks, and every clawback gets reviewed by the Court, is what they're saying.

THE COURT: Number one, I'm not going to do that. But because this number of documents is small, because we've

P8CEOPE4

1 encountered some privilege issues already today, I'm inclined
2 to rule the same way that I did for the other privilege issues.
3 So you don't need to make any arguments about the 18 docs, but
4 let's hear from you about the five documents, but understand
5 where I'm probably going with this.

6 MR. GRATZ: I understand, your Honor. And I think
7 there is a distinction between the five and the 18, because
8 they are saying there is a problem with the five, and all they
9 are saying about the 18 is that may be totally privileged, but
10 you clawed it back, so it should be subject to *in-camera*
11 review.

12 THE COURT: Okay. I'm already trying to put that
13 aside because I am willing to review 23 documents under the
14 same timing and proposal or order that I had directed earlier,
15 which is you get a chance to see if you can narrow the number
16 of documents or whether you want to keep them up, but I'm going
17 to review them and loser pays. And if some of them are
18 privileged and some of them are not, guess what, I'll figure
19 out a way to apportion costs.

20 MR. GRATZ: Understood, your Honor.

21 Let me address the five documents. And let me
22 suggest, your Honor, that this should be limited to those five
23 documents, and the reason is the parties have joined issue on
24 the question of whether they have privileged. And those
25 documents are documents that we acknowledge, on their face,

P8CEOPE4

1 other than the redacted part, are about business, and that is
2 why we didn't redact the part about business. Similar to my
3 friend's argument that, well, these documents that you withdraw
4 your privilege claims on don't look like very strong privilege
5 claims, and that's why we withdrew the privilege claims, your
6 Honor. Even in the face of the plaintiffs not being willing to
7 expressly, at least, agree that they wouldn't argue that we
8 were thereby waiving.

9 So, my points with respect to the five, in particular,
10 your Honor, is, documents that are about business that then, in
11 the middle of them, recount legal advice shouldn't be held in
12 their entirety but they should be redacted, and that describes
13 these documents.

14 THE COURT: So the five documents at issue in the
15 first category are really about the redactions.

16 MR. GRATZ: I think that's right, your Honor.

17 THE COURT: Yes.

18 Mr. Connors?

19 MR. GRATZ: I don't know if my friend would agree.

20 MR. CONNERS: We don't believe that they are
21 privileged. We have a good-faith belief that they are not
22 privileged. And yes, they have been redacted, I think all
23 five, and so the question is whether those --

24 THE COURT: Redactions.

25 MR. CONNERS: -- redactions reflected confidential

P8CEOPE4

1 legal advice or not.

2 THE COURT: Okay. So you'll get a short period of
3 time to see whether you want to withdraw your requests or
4 narrow the number, reduce the number of documents that I'm
5 going to be reviewing *in-camera*.

6 (Continued on next page)

P8CJOPE5

1 THE COURT: And then you're going to -- if it's --

2 MR. GRATZ: And your Honor, with respect to these
3 documents at issue, we submitted declarations discussing them
4 that support the privilege claim. And we -- we think that that
5 is -- that goes well above and beyond.

6 THE COURT: So now you don't want an in camera review?

7 MR. GRATZ: We do not want in camera review.

8 THE COURT: But if they persist and I review and I
9 agree with you, who has to pay? Who is the loser in that
10 scenario? We'd argue we're all losers in some form or
11 another --

12 MR. GRATZ: I think we all, including the Court, are
13 the losers -- would result from unnecessary in camera review.
14 We think what the Court should do is look at the declarations
15 we have provided with respect to all these documents and
16 conclude that there's no justification for in camera review
17 here. Or at most, review in camera five documents that we've
18 lodged with the Court for in camera review, and the Court can
19 make a determination on those that may well provide guidance.

20 THE COURT: Nope. File a joint letter by August 22
21 whether you've resolved this dispute or narrowed it to some
22 number of documents fewer than the 23. Tell me what the
23 outcome is, whether you still want in camera review of a
24 certain number of documents, which I think is up to the
25 plaintiffs unless you're agreeing that some subset of the 23

P8CJOPE5

1 can be produced in the form that they seek.

2 So you're going to tell me what's still at issue on
3 August 22, and I will undertake an in camera review of the
4 disputed documents. You're going to give me a proposed
5 schedule. Again, I want it to be on a pretty short timeline,
6 but I have my September motion list coming up, so I want to
7 make sure that I don't get stuck, but you are permitted to
8 certainly submit some letter briefs that talk to me about the
9 particular documents that are still at issue. And I will
10 review them in camera, and I will apportion costs or assess
11 costs under 37(a)(5).

12 MR. GRATZ: One request that we would make of your
13 Honor in order to encourage here and perhaps elsewhere the
14 parties to resolve rather than maintain disputes, that the
15 production of documents that is backing down from a privilege
16 position does not operate as a waiver outside --

17 THE COURT: This is specific documents.

18 MR. GRATZ: -- outside the four corners of the
19 specific document.

20 THE COURT: Agreed.

21 MR. CONNERS: May I be heard briefly, your Honor?

22 THE COURT: Why if I'm going to be looking at
23 documents? This is something you should be agreeing to.

24 MR. CONNERS: Here is the problem. We have a concern
25 about selective production of documents where they might

P8CJOPE5

1 produce -- let's say it's a legal issue like let's say it's a
2 fair use defense where they might produce documents that go one
3 way about the fair use defense, but not documents that go
4 another way, and it's very hard for us to even probe into that
5 issue based on the privilege logs that we have, which don't
6 give that level of specificity, but that's the concern. That's
7 the reason why us foregoing our ability to argue waiver
8 prejudices us, and we don't think that should be a condition of
9 them producing documents that aren't privileged in the first
10 place.

11 THE COURT: No, no, no, no, no, no. It's not a
12 condition. It's not a condition, okay? If you can agree to
13 produce documents because they are or are not privileged,
14 that's it. Going forward, if there are other documents that
15 you're concerned about, we'll take them up again. And maybe
16 after a few periods of assessing costs we'll figure out, okay.

17 MR. CONNERS: Right. Thank you.

18 THE COURT: If you're concerned about selective
19 waiver, the answer is to bring up other issues, to bring up
20 other potentially privileged issues for in camera review. I'm
21 not suggesting that you do that because it takes a lot out of
22 my time. It's a lot of what I'm doing here. But this is the
23 first time I've heard you all saying that you don't want in
24 camera review, which actually makes me even more curious about
25 what these documents are about.

P8CJOPE5

1 MR. GRATZ: We have provided these documents for in
2 camera review so your Honor can look at them. We don't think
3 that in camera review is necessary. We are the proponents of
4 the privilege here. They are the opponents of the privilege
5 here.

6 THE COURT: Have your discussion, produce as you make
7 your decisions to produce. That's not up to me, right? I'm
8 not going to -- but I'm not going to add conditions to that
9 production, okay? If there is a concern about selective
10 production, if there's a concern about other documents that are
11 currently being withheld that may be privileged, I may regret
12 saying this, but we'll deal with them next time.

13 MR. GRATZ: Here is the problem, your Honor, which is
14 when we have made a privilege claim and they're saying that
15 privilege claim is no good, one of the ways to resolve that in
16 meet and confer is for us to decide, you know, in context we
17 think we're right, but this document isn't very important and
18 we're just willing to produce it to resolve the issue. If by
19 doing so we are walking into a potentially broader privilege
20 waiver, we're not going to be able as readily to do that, and
21 there will be more -- that will harden the parties' respective
22 positions rather than promoting cooperation and sort of the
23 resolution of disputes. That's the thing that I'm concerned
24 about.

25 THE COURT: I hear you. I hear you. You're going to

P8CJOPE5

1 talk about this. You're going to figure it out. I don't think
2 23 documents is going to kill me.

3 MR. CONNERS: Your Honor, I will note I'm not quite
4 sure how the process of awarding costs to the prevailing party
5 is going to work, but I would note on these documents, the
6 ten-document sample, I feel we are already the prevailing
7 party. We filed a motion.

8 THE COURT: No, no, no. It's five documents now.
9 You've worked out the other stuff. You've got five documents
10 left.

11 MR. CONNERS: Well, we had to file a motion and
12 then --

13 THE COURT: I know. Look, there's so much heat here.
14 Y'all do a really good job of trying to tone it down and tamp
15 it down. But it's late, everybody's cranky, nobody's eaten,
16 and I've had to put off some of my conferences for the
17 afternoon to try to finish out what we're trying to do here,
18 okay? So you're all very experienced and capable lawyers. You
19 will figure out a way either to resolve it or to narrow the
20 issues or give me all 23.

21 MR. CONNERS: Thank you, your Honor.

22 THE COURT: Okay.

23 MR. GRATZ: Just to be clear, your Honor, as to the
24 ones where they are clawbacks, your Honor is ordering in camera
25 review, we think those are all privileged. If they're all

P8CJOPE5

1 privileged, we get fees on those.

2 THE COURT: Right. Or they can back down from that
3 and it's not an issue, right. Because what I've heard on the
4 argument on the clawbacks, which I agree with you, is that
5 there's a right to have in camera review. That may be, but I'm
6 saying that right comes with a slight risk, right? Take a look
7 and see whether this is what you really want to fight about,
8 and we'll look at Rule 37(a)(5) at the same time.

9 MS. BROOK: Moving us forward, docket 429 -- sorry,
10 docket 393. The opposition is docket 429.

11 THE COURT: Oh, this is the deposition of the five
12 former class plaintiffs. So before you start, class cert
13 hasn't happened yet, right?

14 MS. NIGHTINGALE DAWSON: That is correct, your Honor.
15 Alana --

16 THE COURT: Why does this matter? That's a signal.

17 MS. NIGHTINGALE DAWSON: Thank you, your Honor. I
18 have received your signals before, so I appreciate the signal
19 here. Alana Nightingale Dawson on behalf of OpenAI.

20 This matters is for a few different reasons, your
21 Honor. Number one, we are seeking, to be very clear,
22 depositions of five individuals regarding issues that the
23 plaintiffs content are common to the class that go directly to
24 the merits of this case as well as to class certification. So
25 this is not specific only to class certification.

P8CJOPE5

1 However, in this context as I understand it, discovery
2 is occurring in tandem for both merits and class certification.
3 With respect to our specific motion, we are seeking narrow,
4 time-limited depositions from plaintiffs who pursued this
5 litigation for almost two years, have unique personal knowledge
6 that is relevant, again, to the issues. And I will point your
7 Honor to the places in the complaint and the allegations that
8 these specific five people have knowledge about and personal
9 experience with that plaintiffs say are common issues that
10 OpenAI cannot otherwise attain.

11 With respect to the standard, I think that Judge
12 Liman's decision in the *Fishon v. Peloton* case is instructive.
13 As the Court discussed there, the concept of two strangers,
14 which is what the absent class standard and discovery rules
15 protect against, these plaintiffs are not true strangers. They
16 are the opposite. And in fact, when you actually look at the
17 brief and you look at plaintiffs' response – and notably, it is
18 just plaintiffs, not the deponents who we seek who are opposing
19 – their response does not claim that there is any burden.
20 Plaintiffs have never claimed burden, and we have gone out of
21 our way to make sure these won't be burdensome. We will limit
22 them to three hours. We will do them in a time and place that
23 works for plaintiffs, and we're willing to do them remotely.

24 The plaintiffs have never claimed that we are seeking
25 this for an improper purpose or to harass or alter the

P8CJOPE5

1 composition of the class – again, factors Courts consider. The
2 only thing that plaintiffs raise in objection is a question of
3 potential relevance and duplicativeness. But even there, your
4 Honor, again, as to both the merits and class cert, both
5 issues, each of these plaintiffs have personal, highly relevant
6 knowledge.

7 And I'd like to make a few examples, if I may. I know
8 we're late, but I think the few examples are relevant, but I
9 wanted to make sure I responded to your Honor's signal before I
10 get to the specific plaintiffs.

11 First, I'd like to point out with respect to Plaintiff
12 Winchester, among other allegations in the complaint,
13 plaintiffs allege in the paragraph 314 – and again, these are
14 allegations that plaintiffs say are common issues – that the
15 class has been and continues to be irreparably injured due to
16 defendants' conduct for which plaintiffs claim there is no
17 adequate remedy at law, which looks at, among other things,
18 whether there can be monetary compensation.

19 Mr. Winchester is the only person amongst – as my
20 understanding – plaintiffs and forme plaintiffs who signed up
21 for the very HarperCollins deal that plaintiffs rely on at
22 paragraph 11 of their complaint when they argument an
23 entitlement to licensing, which is contrary to the notion that
24 there is no adequate remedy at law.

25 Plaintiff Preston is a founding member, an investor in

P8CJOPE5

1 Created by Humans. Plaintiffs allege, again as a common issue
2 that will come up on the merits and in the context of class
3 certification, that whether the conduct at issue here
4 constituted direct infringement is a common fact and a common
5 legal question.

6 Setting aside whether we agree, relevant to that
7 question is our fair use defense. And Created by Humans,
8 plaintiffs say, is an AI licensing platform that they have
9 pointed out that Plaintiff Authors Guild is partnering with.
10 Plaintiff Preston is identified on their website as having
11 received revenue from that platform. He is also quoted on that
12 platform, and he did, as plaintiff who led this case, did
13 interviews about that platform with the New York Times along
14 with plaintiffs. You can see those representations about his
15 income at 397-5.

16 Plaintiffs don't contend that that's irrelevant. And
17 notably as to Winchester, plaintiffs don't content that his
18 deal is not irrelevant. In fact, they intend to, as we've
19 seen, point to that deal. What they say about the Created by
20 Humans is they simply say, well, the representation on the
21 website is not true.

22 We of course can't rely on plaintiffs' letter to the
23 Court as evidence to determine whether or not the
24 representation is true, whether Mr. Preston has received this
25 revenue that Created by Humans promotes him as receiving. We

P8CJOPE5

1 need to ask him about that. In addition, he helped lead this
2 lawsuit. He is the first signatory on a letter that plaintiffs
3 rely on at note 52 and 54. And without getting into specifics
4 because it deals with documents that plaintiffs have designated
5 confidential, he has said in writing, in communications that
6 have been produced that no other party is on, things that are
7 contrary to allegations in the complaint regarding the claims
8 plaintiffs are making regarding harm to their profession.

9 With respect to Silverman, Coates, and Diaz – and
10 again, it's late, so I'll just make one high-level point and
11 can provide specifics if requested – plaintiffs' primary
12 argument is that they were already deposed about the matters
13 that we want to ask them about in the *Kadrey v. Meta*
14 litigation. And they say in fact, the lawyers from Meta asked
15 many questions about a *Daily Show* segment that Ms. Silverman
16 engaged in. During that segment, Ms. Silverman talks
17 specifically about this lawsuit, specifically about OpenAI, and
18 specifically about ChatGPT. The plaintiffs nowhere explain how
19 OpenAI or Microsoft could use transcripts from another
20 litigation that we're not parties to to defend itself in this
21 case as evidence at summary judgment or trial.

22 As Judge Liman noted in the *Fishon* decision, a
23 defendant has a due process right to defend itself. And so
24 while we certainly are mindful of limiting any burdens, there
25 is no claim of burden here, and plaintiffs do not dispute that

P8CJOPE5

1 the deposition transcripts and what those plaintiffs have to
2 say is relevant and that Mr. Preston's role in this case has
3 relevance, including with Created by Humans, or that
4 Mr. Winchester's deal is relevant.

5 Given the highly relevant nature of discovery, the
6 lack of any burden, and the proper purpose for which we seek
7 these, we request an order for limited, not to exceed
8 three-hour depositions. If the Court doesn't have any
9 questions --

10 THE COURT: So is Junot Diaz no longer an issue?

11 MS. NIGHTINGALE DAWSON: So there are other points I
12 could make. So there's Silverman, Coates, and Diaz. So Diaz I
13 think actually is in illustrative example. For Diaz,
14 plaintiffs say, well, you're misrepresenting his testimony.
15 Respectfully, your Honor, I would, A, encourage the Court to
16 look at the docket. It is 339-5. But I think this is an
17 illustrative point, which is that plaintiffs say we are
18 misrepresenting. In fact, we seemingly have different
19 interpretations of his testimony. But importantly, that is
20 testimony that occurred in a different case without any of the
21 lawyers present in this courtroom, to my knowledge. Certainly
22 none of defendants' counsel was there.

23 It was testimony about OpenAI and about ChatGPT,
24 directly relevant to the common allegations that plaintiffs
25 make about our products and us as a defendant. And we have a

P8CJOPE5

1 due process right to ask Mr. Diaz about those statements. And
2 plaintiffs should not be permitted to drop 27 plaintiffs from
3 this case and then shield entirely from discovery the
4 information that is unique and specific to those.

5 Again, we're not seeking depositions from absent class
6 members writ large. We're not seen seeking depositions from
7 all 27 of the plaintiffs that were dropped from this case
8 without any notice to us. What we are seeking is limited,
9 narrow, targeted depositions on matters that frankly I do not
10 read plaintiffs' papers as disputing are relevant. At one
11 point they say minimally relevant. At another point they say
12 we can use their transcripts.

13 But again, that as far as I'm aware is not admissible
14 evidence that we can rely on here, and that is what we are
15 seeking in furtherance of our due process right to defend
16 ourselves.

17 THE COURT: All right. Let's hear from plaintiffs'
18 counsel.

19 MS. SALINAS: So I think we just need to take a brief
20 step back. Why are we having this dispute right now? Because
21 last year OpenAI got up and said we need an MDL. OpenAI could
22 have deposed all these people. The cases remained on separate
23 tracks. But they insisted and got what they wanted, which is
24 an MDL and a consolidation of the class case.

25 Given that consolidation, class counsel had every good

P8CJOPE5

1 reason to create efficiencies and limit the class
2 representatives as we have done. This is a result of their own
3 making, and they don't like the result, but the notion that we
4 are somehow trying to hide our clients from discovery is
5 absolutely false. The only reason we are here is because
6 OpenAI sought an MDL and consolidated the two class cases.

7 The second thing I will note is I think your Honor had
8 it spot on. Theses are issues that would pertain in class
9 certification, and OpenAI again insisted that class
10 certification come after summary judgment. The only issue
11 before the Court on summary judgment is going to be the class
12 representatives' statements, their understanding and whatnot.

13 They want to litigate -- they want to go past that.
14 Again, they got their summary judgment before class
15 certification. They don't like that in this instance, and they
16 want to jump ahead and get to discovery, ask we have no reason
17 to believe will actually be needed until you depose all the
18 class representatives.

19 Now, let's talk about some of the arguments they made
20 for the individuals. Let's start with Diaz, Silverman and
21 Ta-Nehisi Coates. Essentially what OpenAI is seeking is to
22 refer to some passing statements these individuals made in an
23 entirely different lawsuit against Meta about OpenAI. And the
24 argument is that, well, these statements are highly relevant,
25 non-duplicative information that pertains to class

P8CJOPE5

1 certification and factor four. That is simply not the case.
2 And all have you to do is look at the Court's decision in
3 *Anthropic* that granted class certification. In that decision,
4 Judge Alsup made very clear that questions of actual harm do
5 not predominate over the bigger issue of the common questions
6 of how the infringement occurred and that that common question,
7 that predominates over any questions of individual harm. They
8 do not dispute that.

9 They have no authority to say that these passing
10 statements by one of what will be hundreds of thousands of
11 class reps warrant a deposition here, particularly given the
12 heavy burden that Courts have repeatedly acknowledged applies
13 for absent class members.

14 Now, regarding Simon Winchester, they say they need
15 this information from Simon Winchester about the HarperCollins
16 deal. There is one other party that that information could
17 clearly be obtained from, and that is their co-defendant,
18 Microsoft. Microsoft is the entity that entered the deal with
19 HarperCollins. Microsoft is going to have all the information
20 about the authors that entered the deal, why they entered the
21 deal, all of that.

22 If, after Microsoft is deposed, and they feel there is
23 still lack of sufficient information, then we can revisit if
24 something from Simon Winchester is needed. But there is no
25 reason to depose, again, the heavy burden requirement to depose

P8CJOPE5

1 an absent class member when the party to the deal that is at
2 issue in this case is already going to be deposed.

3 And then when it comes to Preston, something that my
4 colleague left out is that all the information they are seeking
5 regarding Created by Humans, regarding the Authors Guild, all
6 that information is going to be obtained from the Authors Guild
7 that remains a party to this case and Mary Rasenberger, who is
8 the current president of the Authors Guild. She is on the
9 advisory board for Created by Humans. The correspondence that
10 they contend is so relevant is something that Mary Rasenberger
11 is going to have knowledge about, and they can serve a 30(b)(6)
12 topic on the Authors Guild asking about this.

13 Again, this is all the information they can get from
14 the folks that remain in the case. There is way that they can
15 meet the heavy burden of obtaining this additional discovery.
16 Happy to answer any questions the Court may have, but I believe
17 we've addressed the five individuals. And I will note counsel
18 misspoke. We are absolutely objecting on behalf of the five
19 former class representatives that they're seeking to depose.
20 It's not true that we're simply doing it as class counsel.

21 THE COURT: All right. I'm going to have to write on
22 this. Do you see a need for supplemental briefing, given what
23 we've discussed today or not?

24 MS. SALINAS: I don't believe so, but if there's a
25 particular question that your Honor has, we're happy to --

P8CJOPE5

1 MS. NIGHTINGALE DAWSON: I don't need a need for
2 supplementary. I think if there's 30 seconds to respond
3 briefly. If your Honor would prefer I not, right now --

4 THE COURT: I'm going to reserve decision anyway, so
5 unless it's something that's not what you discussed or not in
6 your writing.

7 MS. NIGHTINGALE DAWSON: I think just one final point.
8 I heard again repeatedly the reference to the heavy burden.
9 And on that in particular I would just point your Honor again
10 to the *Peloton* decision and Judge Liman's statement that if
11 there is no improper purpose -- and it is, as I understand,
12 undisputed that there's no proper purpose -- if the burden on
13 the third party is not particularly great, for example, if the
14 third party -- and I'm quoting -- is well healed or already has
15 counsel, again, as is the case here, we just heard that counsel
16 is representing the former plaintiffs, the showing of necessity
17 is not particularly strong." And that is at 336 F.R.D. at 71.

18 We thank your Honor for taking it under advisement.
19 Obviously have specific responses on the specific issues, but I
20 trust the papers will address this.

21 THE COURT: Are we on our last issue then?

22 MS. BROOK: Your Honor, last issue, docket 413.

23 THE COURT: Oh, this issue.

24 MR. SMYSER: Craig Smyser for the class plaintiffs.

25 Over the past 15 months, OpenAI has changed its

P8CJOPE5

1 position on the reasons for the deletion of the pirated books
2 one and books two data sets five times. First, 15 months ago,
3 when briefing a discovery motion, OpenAI told plaintiffs'
4 counsel and this Court that books one and books two, sourced
5 from LibGen and at the heart of this case, were deleted due to
6 "nonuse." That's docket 145 at page 3 and docket 143 Exhibit D
7 at two.

8 Second, eight months ago, when trying to prevent
9 plaintiffs from deposing a witness on the reasons for the
10 deletion, OpenAI claimed privilege. That's docket 362 at one.
11 Third, just over two months ago in front of your Honor, when
12 trying to prevent the Court from finding waiver or that the
13 crime-fraud exception to attorney-client privilege applied,
14 counsel for OpenAI told this Court that OpenAI is, quote, not
15 asserting privilege and would not block, quote, plaintiffs from
16 exploring the question of nonuse of the data sets to be the
17 cause of the deletions. That's the morning session of the
18 May 2025 hearing, 71, 11 through 14.

19 Fourth, a month ago, OpenAI's position was that it
20 would not rely on any non-privileged reasons for the deletions
21 of the data sets in this litigation. That's exhibit 10 to the
22 motion at page 7. And fifth, finally, approximately 12 days
23 ago and only after plaintiffs took the deposition which this
24 Court ordered to determine the extent of the privileged claims
25 regarding the deletion of books one and books two, OpenAI's

P8CJOPE5

1 position has come momentarily to rest on all the reasons are
2 privileged. That's exhibit 11 at one.

3 The court should not approve this conduct and should
4 require production of the purportedly privileged material
5 underlying the reasons for the deletion. OpenAI -- and that's
6 for two reasons. First, is the at-issue or implied waiver
7 argument. OpenAI has opportunistically used nonuse and
8 privilege implications to leave plaintiffs and the Court to
9 pursue discovery into the reason for the deletion of these data
10 sets for more than a year, including in filing multiple
11 motions, spending hours deferring, not to mention hours at the
12 depositions themselves. Yet OpenAI is using privilege to block
13 discovery.

14 This is a classic sword and shield issue. Plaintiffs
15 did not force OpenAI to put this at issue. OpenAI did that.
16 And the prejudice that plaintiffs have experienced alone in
17 pursuing this line of discovery justifies a finding of waiver
18 here. As the Court has already been told and has recognized
19 throughout this hearing, the time for discovery is rapidly
20 closing. We've had to spend a year that we cannot get back
21 trying to chase this down. That prejudice alone justifies a
22 finding of implied waiver.

23 It's also at issue due to OpenAI's claim that it acted
24 innocently, not willfully, with an innocent state of mind.
25 *Arista Records* case cited in our papers is very clear on this,

P8CJOPE5

1 as is *Bowne of New York City v. AmBase Corp.* Those are both
2 cited in our papers. I won't waste time quoting them here.

3 Third, I will note it's also at issue in that OpenAI's
4 arguments regarding noninfringement cannot be assessed because
5 we don't have the data sets used to train GPT 3.5 because they
6 deleted them. And we also don't have the full suite of files
7 that they actually downloaded from LibGen, which ultimately
8 form the basis of the GPT-3 and 3.5 datasets.

9 There's a second independent reason justifying the
10 production of the purportedly privileged material here, and
11 that's the crime-fraud exception. OpenAI knew, as was shown in
12 the deposition transcript portions cited in our brief, that
13 there was a, quote, substantial question regarding the legality
14 of training on copyrighted materials. They knew that these
15 data sets in particular were particularly bad on that front, as
16 the quote in the first paragraph of our motion indicates. They
17 involved their litigation counsel, including counsel who have
18 been before you today at the podium, and deleted them at their
19 instruction. These are the only two sets so ever deleted.
20 That is probable cause of litigation conduct which undermines
21 the adversary system. Your honor is more familiar than I am
22 with the probable cause standard.

23 Lastly, criminal copyright infringement. OpenAI
24 argues that this would pre-decide the merits, but that's not
25 the case. Fair use is an affirmative defense. That a

P8CJOPE5

1 defendant has an affirmative defense is not in any way a bar to
2 finding probable cause of a crime. I welcome the Court's
3 questions.

4 THE COURT: Let's hear from the defendants.

5 MR. SLAUGHTER: Good afternoon, your Honor. Much
6 later in the afternoon than I think any of us would have liked
7 to be, but thank you very much for your time and patience
8 today. James Slaughter on behalf of OpenAI. I'll respond.

9 Your Honor, OpenAI has been consistent entirely that
10 the reasons for the removal of those data sets are privileged.
11 And we have provided plaintiffs all of the relevant discovery
12 and facts they need with respect to books one and books two
13 datasets. And let me just refresh the Court's recollection of
14 what has happened already and what's been produced and the
15 discovery they've taken, because they've taken all of the
16 discovery they possibly could need with respect to books one
17 and books two, and here's where we are.

18 No ChatGPT model currently available has been trained
19 on books one and books two datasets GPT-3.5 was the last model
20 trained on those data sets and was released in 2022 and was no
21 longer used after early 2023. Those datasets were removed in
22 2022, a full year before any plaintiff filed any litigation
23 challenging OpenAI's training on books. And we have since
24 recovered and produced those datasets to the plaintiffs. And I
25 want to emphasize that. They have this data and they can do

P8CJOPE5

1 their analysis.

2 And so plaintiffs now seek discovery not into the
3 content of those datasets, not into when they were used or when
4 they stopped being used, because they have all of that
5 information. They're now seeking discovery and continue to
6 seek discovery into the reasons for their removal. But OpenAI
7 has consistently explained that the reasons for the removal are
8 privileged because they were based upon advice of counsel.

9 OpenAI has informed plaintiffs that there are no
10 non-privileged reasons for the removal, and we will not be
11 advancing any. And we're not relying on the advice of counsel,
12 either affirmatively or in rebuttal, in response to any claims.
13 We are not putting this as issue. It's plaintiffs that are
14 trying to do so.

15 And so there's been extensive discovery. You've
16 ordered some 30(b)(6) depositions. And at those depositions,
17 there was testimony about when the books were downloaded, who
18 was involved in that downloading, how it was downloaded. There
19 was testimony with respect to the removal, about which
20 attorneys were involved in the removal, when those discussions
21 took place, the subject matter of those discussions, how the
22 removal occurred, who removed them. There has been a great,
23 great deal of discovery into all of those issues about removal.

24 In other words, plaintiffs have all of the
25 non-privileged information available about when the datasets

P8CJOPE5

1 were downloaded, by whom, and the people involved, and the
2 privileged nature of the discussions about their removal. So
3 we're left, your Honor -- and by the way, we're left, your
4 Honor, with the question about waiver. They're now suggesting
5 that there's been some sort of waiver. There hasn't been a
6 waiver, your Honor.

7 All of the cases -- we've told the -- you know,
8 there's been discussion about sword and shield. We are not
9 relying on the advice of counsel. That's what the sword and
10 shield principle is all about. We've told the plaintiffs
11 repeatedly that OpenAI does not intend to introduce evidence of
12 the advice of counsel either affirmatively or in rebuttal to
13 plaintiffs' burden on any claim.

14 We are not putting this at issue. Plaintiffs are
15 attempting to put it at issue. And we've consistently said
16 that the reasons for this removal are privileged because
17 they're based on legal advice. And it's relevant, your Honor,
18 I think, and helpful to look at the cases on waiver because
19 this is where we're getting. And all of the cases about waiver
20 discuss reliance. There has to be an element that the party
21 asserting the privilege is relying on privileged communications
22 in part to support a claim or defense.

23 And that's obviously not at all what is going on here.
24 It's just the opposite. And I would refer your Honor to the
25 most recent decision we found is the *GLD3, LLC v. Albra* case in

P8CJOPE5

1 this district from 2024. And I think that the language in that
2 case is --

3 THE COURT: Is there a docket number for that case?

4 MR. SLAUGHTER: That case, *GLD3 v. Albra*, is 2024 WL
5 4471672. And that case quotes the Second Circuit in *In re*
6 *County of Erie* from 2008 which emphasizes that a privilege
7 waiver only happens when a client places the attorney-client
8 relationship directly at issue. So the idea would be somehow
9 that OpenAI is putting this quote directly at issue and asserts
10 reliance on attorney's advice as an element for a claim or
11 defense.

12 That could not be further from the truth. We have
13 been steadfast and consistent that we are not relying in any
14 shape or form on advice of counsel with respect to the reasons
15 for the removal of datasets for any claim. And that's the --
16 and that is really -- and the real sort of dichotomy or -- I'm
17 missing the word, but there's sort of two different issues.
18 One is this removal of these datasets and subsequent recovery
19 and production to the plaintiffs so they have the data.

20 But there's a separate question about use, right?
21 Okay. So they were used for a period of time in training and
22 fair use, and there will be loads of evidence on fair use. And
23 so -- but the reasons --

24 THE COURT: Yes. So I guess you were about to say,
25 but why were they removed?

P8CJOPE5

1 MR. SLAUGHTER: The reasons for the removal are all
2 privileged, your Honor. We've said to the plaintiffs --

3 THE COURT: But you're not relying on advice of
4 counsel? They're privileged reasons because presumably they
5 were removed -- I'm trying to understand how this isn't
6 circular. The pirated datasets were removed, but we can't tell
7 you why because the reason is privileged, but we're not relying
8 on advice of counsel, but counsel told us that we should remove
9 them.

10 MR. SLAUGHTER: We're not -- we are not offering any
11 evidence, your Honor, about the removal of datasets. We're not
12 offering them on any subject. The plaintiffs had asked us why
13 did you remove them, and we said the reasons we've removed them
14 are privileged. We are not advancing -- we're not relying --
15 we're going to go to trial. We're going to go to summary
16 judgment. We're not advancing the reasons for removal or
17 discussing removal.

18 If plaintiffs are going to want to discuss it, they
19 have all they need to discuss it. They can say they downloaded
20 on such and such a date, they used it to train such and such a
21 models, they subsequently removed it, and so you can draw the
22 conclusions you want to draw from it.

23 THE COURT: So then the conclusion would be that there
24 would be -- there was some understanding or knowledge that
25 maybe it was not really a great idea to use these datasets, so

P8CJOPE5

1 we're going to remove them, but we're not going to tell you
2 that that's why it was. I can't ask you to respond to that,
3 but doesn't this go to knowledge or awareness, which is
4 something that the plaintiffs need to prove?

5 MR. SLAUGHTER: I don't know if it is something that
6 the plaintiffs need to prove, your Honor. But we are not
7 relying on it for any claim or defense that we have to bear the
8 burden on. So --

9 THE COURT: Well, but that's only half the issue,
10 right? I mean you're not relying on a defense, yes, because
11 maybe it sucks.

12 MR. SLAUGHTER: Your Honor, we're not relying --
13 there's --

14 THE COURT: So there's no reason. There's no reason
15 why --

16 MR. SLAUGHTER: No.

17 THE COURT: -- a fact finder would then be permitted
18 to draw an inference from the lack of evidence because the
19 evidence is being withheld?

20 MR. SLAUGHTER: The fact finder can draw whatever
21 inference that the fact finder wants to draw from the
22 non-privileged evidence. And the non-privileged evidence is it
23 was acquired and downloaded on such and such a date, and was
24 used to train such and such a model for whatever period of
25 time. It was subsequently removed and recovered. And that's

P8CJOPE5

1 the -- what they can rely upon, and that's the non-privileged
2 information they can rely on --

3 THE COURT: Wait, wait, wait. So there is an
4 admission though that that data set was used to train GPT-3.5?

5 MR. SLAUGHTER: Yes.

6 THE COURT: Okay. So does that mean that there's an
7 acknowledgment or an admission that infringing material was
8 used to train GPT models that are at issue in this case?

9 MR. SLAUGHTER: We -- if the question is whether there
10 was copyrighted material used to train models, that's the --

11 THE COURT: Copyrighted material that was obtained
12 with or without knowledge, that it was obtained from a pirated
13 database? I mean, I think one of the issues in this particular
14 dispute is that at some point in time somebody figured out that
15 these were not -- they were obtained from a not-so-great
16 source.

17 MR. SLAUGHTER: That's a different question than the
18 question that's on this motion. The question on the motion is
19 the reasons -- the waiver based upon the reasons for removal.
20 You should tell us what the reasons are for the removal, and we
21 said all along that they are privileged.

22 The question that you're asking is a different
23 question, which is when did you download, which we've given,
24 and what did you know about them? They can ask all of those
25 questions about when did you download them and what did you

P8CJOPE5

1 know about them at the time.

2 THE COURT: So if a lawyer -- hypothetically, of
3 course, because I haven't looked at the discovery and I'm
4 not -- so hypothetically, if someone were able to take a
5 deposition and say, okay, you downloaded these, and at what
6 point in time did you come to find that these were acquired
7 from LibGen? What do you know about LibGen and when and who
8 told you? Doesn't that get you right --

9 MR. SLAUGHTER: I think that the -- but the question
10 on this motion is the reasons for the removal, and so you
11 can -- there can be questions at a deposition. You acquired
12 this from LibGen? Yes, I did. When did you do it? How did
13 you do it? What did you know about it?

14 THE COURT: Take me through every step of your
15 knowledge.

16 MR. SLAUGHTER: Yeah. And I think when they get to
17 the question about was it removed? Yes. Why was it removed?
18 Those answers are privileged, and we've always said they're
19 privileged. And we've said that there are no non-privileged
20 bases for the reasons for the removal. So this is just this
21 point at the very end about why was -- were these datasets
22 removed, and then obviously they were subsequently recovered
23 and produced.

24 But the question on this motion is why were they
25 removed, and that question is entirely privileged. The answer

P8CJOPE5

1 to that question is entirely privileged. We've been consistent
2 about that all along. And we've explained to plaintiffs that
3 there are no non-privileged reasons for the removal.

4 THE COURT: All right. I see Mr. Smyser shaking his
5 head. Come on up and talk to me.

6 MR. SMYSER: Your Honor, there's a key point here that
7 I think is crucial to understand. Mr. Slaughter repeatedly
8 asserted something that is not true. They have not recovered
9 the datasets used to train GPT-3.5. They have also not
10 recovered the original downloads from LibGen. Those downloads
11 are in this case. They are part of how the datasets were
12 created. We cannot determine what works were included in them.
13 They are, as Judge Alsup recognized in the *Anthropic* case,
14 themselves an independent act of infringement. We also cannot
15 determine because of their deletion the contents of the GPT-3.5
16 training datasets because they no longer exist because they
17 were removed.

18 Secondly, he repeatedly stated that OpenAI has been
19 consistent that these reasons were all privileged. Search your
20 memory. You were at the May hearing. You heard their counsel
21 say that all the reasons were not privileged. They've changed
22 their position on this multiple times because they originally
23 put in a non-privileged reason into the record as part of an
24 argument to this Court. That is sufficient on its own to put
25 it at issue.

P8CJOPE5

1 But even if you disagree with that, and then the
2 shifting positions which have caused us massive prejudice over
3 the past 15 months, even if you disagree with that, there is no
4 question that the reasons they deleted these datasets and
5 removed them – again, the only ones they have ever done so –
6 goes directly to OpenAI's state of mind, which is what your
7 Honor was touching on.

8 This is the *Arista Records* case again cited in our
9 papers. You did not need to be advancing a specific
10 affirmative defense of advice of counsel in order for the state
11 of mind to be at issue. They are arguing that they had an
12 innocent, good-faith belief in the legality of their actions in
13 order to contradict a finding of willfulness, which goes
14 directly to a statutory damages enhancement.

15 They'll be arguing that to the jury. We need the
16 materials in order to test that. It's a classic sword and
17 shield issue. And the unfairness that plaintiffs have frankly
18 already suffered at the hands of OpenAI's conduct is sufficient
19 under the implied waiver doctrine to find a waiver here.

20 A few more minor points. I believe Mr. Slaughter said
21 that they've allowed us to ask deposition questions to all of
22 these issues. At the most recent deposition they prevented the
23 witness from answering a yes/no question on whether or not
24 non-privileged reasons for the existence of books one and books
25 two being deleted existed at all.

P8CJOPE5

1 The grounds for that instruction were entirely
2 improper. They were that, witness, I instruct you not to
3 answer because plaintiffs will use that to argue waiver.
4 That's not allowing the witness to answer the question, and
5 it's not a proper privilege instruction. Because at the time
6 of the deposition, their position was not that there are no
7 non-privileged reasons. Their position was we won't advance
8 any non-privileged reasons. That was I think the fourth of
9 their five metamorphoses over the past year.

10 I will also note that the *GLD v. Albra* case is much
11 weaker than they're stating it. Again, I invite your Honor to
12 review *Bilzerian*, *Chipotle*, *Arista*, and *Grund* from our papers,
13 which clearly indicate that there is no requirement that the
14 reliance be expressed or particularly invoked. The reason for
15 that is that that would be express waiver. We're talking about
16 implied waiver where the content of the arguments that the
17 party is advancing necessarily implicate a privileged matter,
18 which would be unfair to allow the other party not examine.
19 That's where we're at here. Couldn't be clearer.

20 Happy to answer any questions.

21 THE COURT: All right. This is another thing I'm not
22 going to rule from the bench. So I think we're winding up
23 here. All right. Mr. Slaughter?

24 MR. SLAUGHTER: I was going to obviously allow you to
25 finish, your Honor, if you had anything more to say. But there

P8CJOPE5

1 were things Mr. Smyser said that I feel deserve a response, but
2 if you --

3 THE COURT: You can respond, and then I'm going to
4 finish up the conference.

5 MR. SLAUGHTER: Thank you so much. Just with respect
6 to the *GLD3*, your Honor, it is directly a case about waiver by
7 implication and it talks about reliance, as I just discussed.

8 Mr. Smyser discussed all the datasets that they say
9 they don't have. A, that's not true. The datasets have been
10 produced to them. B, of course 2022, at the time the datasets
11 were removed, they were removed a year before any litigation
12 had been threatened about training, so there was no duty at
13 that time to retain. And I just want to --

14 THE COURT: Wait. But you just said the datasets were
15 produced.

16 MR. SLAUGHTER: Yes, we recovered them.

17 THE COURT: And you're also arguing there was no duty
18 to retain the datasets that were removed?

19 MR. SLAUGHTER: Correct.

20 THE COURT: What datasets have been produced?

21 MR. SLAUGHTER: We recovered the books one and books
22 two datasets after this litigation was filed, and we searched
23 for them and found them and produced them.

24 THE COURT: Okay.

25 MR. SLAUGHTER: Okay. And then your Honor, just with

P8CJOPE5

1 respect to this point that they have -- and this is in our
2 brief. We -- if your Honor's interested, there was a letter
3 that was written, meet and confer letter and one letter that
4 was written to the Court, and Mr. Smyser characterizes that
5 letter as the purpose of the letter was to describe the reasons
6 for the removal of those datasets.

7 I encourage your Honor to look at that letter. That's
8 not what the purpose of that letter was. The intent of the
9 language in those letters could have been more precise than
10 they were, but it was clear that they were intended to disclose
11 that the data sets had been removed. And at the time they
12 were prior to any litigation being filed, and that OpenAI was
13 actively investigating whether they could be located. And
14 that's what those letter were about. They were not some
15 letters that were intended to describe all the reasons for the
16 removal.

17 Your Honor, Mr. Smyser touched on the crime-fraud --
18 this isn't crime-fraud. This is fair use. As Judge Chhabria
19 found in the *Kadrey* case where he found that it was fair use to
20 train on books that had been downloaded from LibGen. So don't
21 want to not address the crime-fraud issue. If your Honor's
22 interested, that's obviously addressed in our papers.

23 THE COURT: I don't need to hear it again.

24 MR. SMYSER: Your Honor, it's not true. It is not
25 true that they have produced the books one and books two

P8CJOPE5

1 datasets. I want to be very clear about this. There was a
2 version of books one and books two used to train GPT-2. They
3 located a copy of that and have produced it. There was other
4 versions of the books one and books two datasets used to train
5 other models including GPT-3.5. They have not found or
6 produced that to us.

7 They have also not found or produced the original
8 downloads from LibGen. Those have all been deleted. We don't
9 have those. We cannot investigate a claim of noninfringement
10 as to those datasets, for example. So it is simply not true to
11 say they have covered and produced everything. That is not
12 true. I had to make that clear because that's; the record of
13 the case so far, and it's crucially important to understanding
14 what's at issue here.

15 THE COURT: All right. August 22 you're going to --
16 plaintiffs are going to provide me a timeline, your timeline on
17 the flip-flopping of the positions, your alleged flip-flopping
18 of the positions. You're going to cite to ECF numbers if
19 they're already here. In other words, I don't need to see them
20 all produced again. Just give me the ECF numbers and I can
21 look at them. All right? And then August 29, OpenAI can
22 respond to that. And then after that, I will see if I need
23 additional briefing or I have additional questions. Okay?

24 MR. SLAUGHTER: Thank you, your Honor.

25 MR. SMYSER: Thank you, your Honor. And the August 22

P8CJOPE5

1 filing, to be clear, is just the timeline, correct?

2 THE COURT: Timeline with citations to ECF filings or
3 if they haven't been filed yet --

4 MR. SMYSER: As exhibits. Understood.

5 THE COURT: As exhibits. And response from OpenAI
6 August 29. Do not file more on this issue until I say so or
7 ask for it. I am taking this very seriously. It's something
8 that's quite concerning. And if I have other questions, I will
9 ask them. All right. Anything else?

10 MR. SLAUGHTER: Your Honor, nothing on that matter. I
11 did want to raise a housekeeping matter, which perhaps this is
12 where you were headed, but at the risk of another month of
13 discovery reports, it does seem to me we should schedule our
14 next conference together. I don't know if you were thinking
15 about that.

16 THE COURT: I have dates for those. All right.
17 Everybody ready? Tuesday, September 23; Wednesday October 29;
18 and, Thursday, December 4.

19 MS. BROOK: Okay. You're not going to like me. If
20 it's just me, then by golly-gee, go forward. But Tuesday,
21 September 23 is Rosh Hashanah, which I do imagine might impact
22 more than me.

23 THE COURT: I should have been aware of that. I'm
24 sorry.

25 MR. SLAUGHTER: Can we do the 30th, your Honor?

P8CJOPE5

1 THE COURT: September 30. Does that get us past Rosh
2 Hashanah?

3 MR. SLAUGHTER: We will all be atoning, your Honor.
4 That's in between Rosh Hashanah and Yom Kippur.

5 THE COURT: When does Yom Kippur start?

6 MR. SLAUGHTER: Starts the evening of the Wednesday of
7 the next day.

8 MS. BROOK: We could even do later in the week of the
9 22nd, your Honor, so folks that observe both days will be
10 unable to travel until the evening of the 23rd. But if we did
11 it on the 26th, that Friday, that would not be conflicted by
12 the holidays.

13 THE COURT: But the 25th also would not be okay
14 because the travel restrictions isn't lifted into the evening?

15 MS. BROOK: That's not ideal, no.

16 THE COURT: What if we did the 25th in the afternoon?
17 Does that help?

18 MS. BROOK: With that one it's probably just me coming
19 from Los Angeles, and so I will work that out if need be.

20 THE COURT: So maybe September 25 in the afternoon.
21 Let me just take a moment also.

22 All right. The reason why I pushed for the 25th in
23 the afternoon rather than putting it over till earlier the next
24 week is because if we need to have another one of our
25 settlement conferences or conversations, I'd like to try to

P8CJOPE5

1 keep Fridays open for that. And that hopefully means that
2 people wouldn't have to stay over the weekend if they're coming
3 in for the conference.

4 So let's for now put it September 25 in the afternoon.
5 Let's start at 2:30. Is there any issue with the other dates?

6 MS. BROOK: Nothing that I see of that kind, your
7 Honor.

8 THE COURT: I mean, they're far enough out that if
9 there are holidays that didn't appear on my calendar or on my
10 radar at the time we set these dates, we can reassess. But for
11 now, September 25 in the afternoon. We're going to start at
12 2:30. And then October 29, December 4. Okay. What's next?

13 MR. SLAUGHTER: We'll all go get our flights.

14 THE COURT: Oh, you're not coming back on Friday?

15 MR. SLAUGHTER: I will not be here on Friday, your
16 Honor.

17 THE COURT: I guess there's only a subset of you.

18 Anything else we need to do at this time then?

19 MS. BROOK: Nothing else from plaintiffs, your Honor.
20 Thank you and the court for staying so late. We appreciate it.

21 THE COURT: Thank you all for your stamina. Anything
22 from defendants?

23 MS. YBARRA: No, your Honor.

24 MS. HURST: No, your Honor. Thank you.

25 THE COURT: Thank you very much. We are adjourned.