UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK
----------x

IN RE: OPENAI, INC. COPYRIGHT INFRINGEMENT LITIGATION.

25 MD 3143

Conference

----------x

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

Before:

HON. ONA T. WANG,
Magistrate Judge

APPEARANCES

SUSMAN GODFREY LLP
Attorneys for Class Plaintiffs and The New York Times Company
BY: DAVIDA BROOK
ROHIT NATH
ADNAN MUTTALIB
ALEJANDRA SALINAS
JORDAN CONNORS
CRAIG SMYSER
CHARLOTTE LEPIC
HENRY WALTER

THE NEW YORK TIMES
Attorneys for Plaintiffs The New York Times Company
BY: DEMETRI BLAISDELL
KAREN CHESLEY

ROTHWELL FIGG ERNST & MANBECK, P.C.
Attorneys for News Plaintiffs
BY: KRISTEN LOGAN
JENNIFER MAISEL

LOEVY & LOEVY
Attorneys for Consolidated Plaintiff The Center for Investigative Reporting, Inc.
BY: MATTHEW TOPIC

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

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

(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

about surveillance and control by the government of its citizens.

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

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

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

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

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

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

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

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

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

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

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

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

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

THE COURT: Okay.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

MS. BROOK: Good morning, your Honor.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

review.

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

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

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

will be burdensome at all.

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

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

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

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

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

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

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

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

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

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

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

three stages at which infringement is alleged to have happened.

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

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

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

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

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

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

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

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

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

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

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

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

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

deposition.

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

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

(Continued on next page)

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

MS. BROOK: That was Microsoft.

THE COURT: Oh, that was Microsoft. Okay.

Go ahead, Ms. Brook.

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

So, beginning --

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

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

THE COURT: Unfortunately. Sorry.

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

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

THE COURT: Yes.

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

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

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

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

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

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

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

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

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

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

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

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

MS. YBARRA: Yes, your Honor.

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

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

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

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

MS. YBARRA: Yes, your Honor.

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

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

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

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

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

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

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

THE COURT: Okay. Anything else?

Mr. Justice.

MR. JUSTICE: Yeah. Just one quick point.

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

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

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

MR. JUSTICE: Yep.

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

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

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

MR. JUSTICE: No, they have to --

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

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

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

need to review everything before it goes out?

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

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

MR. JUSTICE: Yeah.

THE COURT: Okay?

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

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

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

is protected.

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

THE COURT: Anybody else have anything to add?

I see Ms. Ybarra standing up.

MS. YBARRA: Michelle Ybarra for OpenAI.

Your Honor, I'll be very brief.

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

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

Thank you, your Honor.

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

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

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

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

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

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

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

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

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

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

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

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

MS. BROOK: All housekeeping.

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

THE COURT: Yes.

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

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

MR. SLAUGHTER: Your Honor, just on Friday.

THE COURT: Yes.

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

THE COURT: Well --

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

happen, but --

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

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

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

MR. SLAUGHTER: Thank you, your Honor.

THE COURT: Okay.

MS. BROOK: And three pages, your Honor?

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

All right. Moving right along.

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

THE COURT: Okay.

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

MS. LOGAN: Yes, correct.

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

disconnect is?

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

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

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

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

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

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

MS. LOGAN: July 8.

THE COURT: All right. Then what happened?

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

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

MS. LOGAN: Yes.

THE COURT: Okay.

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

THE COURT: And that was served on July 8.

MS. LOGAN: Correct.

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

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

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

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

MS. LOGAN: Correct.

THE COURT: Okay. So you need a date.

MS. LOGAN: Yes.

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

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

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

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

When can you get this done?

MS. BEYER: Good morning, your Honor.

Carrie Beyer on behalf of Microsoft.

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

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

for --

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

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

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

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

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

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

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

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

Help me out, guys.

MS. LOGAN: September 12.

MS. BROOK: September 12 for substantial completion.

THE COURT: Of document discovery.

MS. BROOK: Of document discovery. Correct.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

be some subcomponent of M365 chat.

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

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

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

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

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

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

MS. HURST: Good morning, your Honor.

Annette Hurst for Microsoft.

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

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

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

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

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

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

THE COURT: Yeah.

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

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

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

And if that's the case and the plaintiffs win on the engine infringing in some way, I think what I heard -- and maybe Ms. Maisel can be clearer -- is, is it really the issue that we're talking about the engine, 4o, or is there something else related to that that may relate to the claims in this case? And I don't know how we get there. You know, because the information is uniquely in the possession of Microsoft

right now.

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

THE COURT: Uh-huh.

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

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

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

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

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

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

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

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

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

Just the big-picture why does this matter, the

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

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

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

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

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

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

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

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

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

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

MS. HURST: Just briefly, your Honor.

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

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

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

MS. MAISEL: So, just one quick point.

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

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

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

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

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

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

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

MS. MAISEL: Thank you, your Honor.

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

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

MS. LOGAN: Thank you, your Honor.

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

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

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

MS. LOGAN: Yes.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

THE COURT: Ms. Logan.

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

infringing outputs coming from the model.

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

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

THE COURT: Okay.

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

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

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

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

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

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

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

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

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

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

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

THE COURT: Anything to add, Ms. Hurst?

MS. HURST: Just briefly, your Honor.

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

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

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

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

MS. HURST: Yes.

THE COURT: Right?

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

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

MS. HURST: Yes, your Honor.

THE COURT: Up and running.

MS. HURST: Right.

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

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

MS. HURST: Correct.

THE COURT: Okay. Anything else?

MS. HURST: No, your Honor.

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

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

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

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

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

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

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

(Continued on next page)

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

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

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

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

But to the extent there are outputs that are some kind

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

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

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

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

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

(Recess)

THE DEPUTY CLERK: Court is back in session.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

can be regurgitated out.

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

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

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

THE COURT: All right. Thank you.

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

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

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

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

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

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

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

THE COURT: Okay. So all right.

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

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

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

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

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

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

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

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

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

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

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

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

MS. SALOMAN: Sure, your Honor.

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

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

THE COURT: Three buckets or three documents?

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

THE COURT: Other presentations have been produced?

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

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

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

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

MR. MUTTALIB: Does the dissemination of that training

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

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

THE COURT: Okay.

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

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

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

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

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

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

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

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

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

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

THE COURT: Oh, yes.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

THE COURT: From 2022 to when?

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

THE COURT: Okay.

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

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

MR. MUTTALIB: July 2023 onwards.

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

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

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

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

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

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

THE COURT: May 2024?

MR. MUTTALIB: That's right.

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

MR. MUTTALIB: December of 2023.

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

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

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

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

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

THE COURT: Okay.

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

THE COURT: Okay.

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

THE COURT: Okay.

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

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

readers, okay. And when was that survey done?

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

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

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

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

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

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

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

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

Honor.

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

MS. NAJEMY: Correct, your Honor.

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

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

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

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

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

THE COURT: And you've gotten the data?

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

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

MS. NAJEMY: Correct, your Honor.

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

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

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

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

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

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

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

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

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

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

THE COURT: You have that information in the survey.

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

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

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

MS. NAJEMY: Thank you, your Honor.

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

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

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

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

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

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

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

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

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

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

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

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

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

the issues out before we get distracted. Go ahead.

(Continued on next page)

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

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

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

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

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

THE COURT: Okay.

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

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

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

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

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

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

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

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

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

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

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

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

THE COURT: Mr. Gratz, come on up.

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

forward.

All right. Mr. Gratz.

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

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

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

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

THE COURT: Okay.

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

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

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

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

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

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

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

MR. GRATZ: So --

MS. BROOK: I have a suggestion.

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

Apply the agreed-upon search terms and any subsequent

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

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

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

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

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

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

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

So, next.

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

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

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

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

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

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

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

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

MR. GRATZ: -- request for production --

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

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

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

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

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

MR. GRATZ: Yes. And --

THE COURT: What about the other three?

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

Ms. Brook's questions got in their entirety.

And I think that simply closes out that issue.

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

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

MR. GRATZ: Yes. We said that.

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

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

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

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

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

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

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

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

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

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

MR. CONNERS: Your Honor, Jordan Conners for class plaintiffs.

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

THE COURT: Okay.

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

they were provided to counsel of record.

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

THE COURT: Thank you. Please do.

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

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

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

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

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

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

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

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

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

MS. BROOK: The issue they are addressing --

MR. GRATZ: I --

MS. BROOK: Oh, sorry. Go ahead.

MR. GRATZ: I'm sorry.

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

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

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

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

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

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

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

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

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

MR. GRATZ: I do.

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

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

MR. GRATZ: An --

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

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

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

MR. GRATZ: Yes.

THE COURT: Your deadline is August 22.

MR. GRATZ: Thank you.

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

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

MR. GRATZ: Yes, your Honor.

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

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

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

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

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

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

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

MR. GRATZ: Two.

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

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

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

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

(Counsel confer)

MS. BROOK: Maroon Shetty.

THE COURT: Shetty?

MS. BROOK: Yes, I believe so.

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

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

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

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

MR. GRATZ: I --

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

MR. GRATZ: Sure.

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

Ms. Brook, what were you going to say?

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

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

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

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

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

MS. BROOK: Yes.

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

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

THE COURT: Okay. Great.

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

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

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

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

MR. GRATZ: Understood, your Honor.

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

MR. GRATZ: Understood, your Honor.

THE COURT: All right.

Next issue.

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

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

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

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

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

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

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

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

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

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

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

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

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

MS. SALINAS: Good afternoon, your Honor.

Alijandra Salinas on behalf of the class plaintiffs.

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

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

THE COURT: So there were ten before?

MS. SALINAS: Yes.

THE COURT: Okay.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

I'm happy to answer any questions.

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

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

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

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

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

THE COURT: One more response.

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

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

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

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

MS. SALINAS: Correct.

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

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

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

THE COURT: Here's what I am going to do:

Plaintiffs, you can go provide your qualified admission to use OpenAI's term. When can you do that by?

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

THE COURT: Okay. August 15.

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

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

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

THE COURT: No.

MS. BROCK: Yeah.

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

it narrows the issues. Okay?

MS. BROCK: Understood.

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

MS. BROCK: Understood.

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

MS. BROCK: Understood.

THE COURT: All right. Next issue.

MS. BROCK: Thank you, your Honor.

MR. CONNERS: Good afternoon, your Honor.

Jordan Conners for the class plaintiffs.

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

THE COURT: There were clawed back; right?

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

So there are two categories: The first category is a

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

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

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

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

Very similar to documents we've seen produced that

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

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

MR. CONNERS: Fair enough.

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

MR. CONNERS: That's exactly right.

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

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

THE COURT: Okay.

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

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

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

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

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

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

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

MR. CONNERS: This is the 18 documents in category

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

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

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

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

MR. GRATZ: Understood, your Honor.

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

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

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

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

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

THE COURT: Yes.

Mr. Conners?

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

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

THE COURT: Redactions.

MR. CONNERS: -- redactions reflected confidential

legal advice or not.

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

(Continued on next page)

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

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

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

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

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

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

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

can be produced in the form that they seek.

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

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

THE COURT: This is specific documents.

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

THE COURT: Agreed.

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

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

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

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

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

MR. CONNERS: Right. Thank you.

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

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

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

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

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

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

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

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

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

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

MR. CONNERS: Thank you, your Honor.

THE COURT: Okay.

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

privileged, we get fees on those.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Plaintiff Preston is a founding member, an investor in

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Given that consolidation, class counsel had every good

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

THE COURT: Are we on our last issue then?

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

THE COURT: Oh, this issue.

MR. SMYSER: Craig Smyser for the class plaintiffs.

Over the past 15 months, OpenAI has changed its

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

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

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

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

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

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

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

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

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

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

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

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

THE COURT: Let's hear from the defendants.

MR. SLAUGHTER: Good afternoon, your Honor. Much later in the afternoon than I think any of us would have liked to be, but thank you very much for your time and patience today. James Slaughter on behalf of OpenAI. I'll respond.

Your Honor, OpenAI has been consistent entirely that the reasons for the removal of those data sets are privileged. And we have provided plaintiffs all of the relevant discovery and facts they need with respect to books one and books two datasets. And let me just refresh the Court's recollection of what has happened already and what's been produced and the discovery they've taken, because they've taken all of the discovery they possibly could need with respect to books one and books two, and here's where we are.

No ChatGPT model currently available has been trained on books one and books two datasets GPT-3.5 was the last model trained on those data sets and was released in 2022 and was no longer used after early 2023. Those datasets were removed in 2022, a full year before any plaintiff filed any litigation challenging OpenAI's training on books. And we have since recovered and produced those datasets to the plaintiffs. And I want to emphasize that. They have this data and they can do

their analysis.

And so plaintiffs now seek discovery not into the content of those datasets, not into when they were used or when they stopped being used, because they have all of that information. They're now seeking discovery and continue to seek discovery into the reasons for their removal. But OpenAI has consistently explained that the reasons for the removal are privileged because they were based upon advice of counsel.

OpenAI has informed plaintiffs that there are no non-privileged reasons for the removal, and we will not be advancing any. And we're not relying on the advice of counsel, either affirmatively or in rebuttal, in response to any claims. We are not putting this as issue. It's plaintiffs that are trying to do so.

And so there's been extensive discovery. You've ordered some 30(b)(6) depositions. And at those depositions, there was testimony about when the books were downloaded, who was involved in that downloading, how it was downloaded. There was testimony with respect to the removal, about which attorneys were involved in the removal, when those discussions took place, the subject matter of those discussions, how the removal occurred, who removed them. There has been a great, great deal of discovery into all of those issues about removal.

In other words, plaintiffs have all of the non-privileged information available about when the datasets

were downloaded, by whom, and the people involved, and the privileged nature of the discussions about their removal. So we're left, your Honor -- and by the way, we're left, your Honor, with the question about waiver. They're now suggesting that there's been some sort of waiver. There hasn't been a waiver, your Honor.

All of the cases -- we've told the -- you know, there's been discussion about sword and shield. We are not relying on the advice of counsel. That's what the sword and shield principle is all about. We've told the plaintiffs repeatedly that OpenAI does not intend to introduce evidence of the advice of counsel either affirmatively or in rebuttal to plaintiffs' burden on any claim.

We are not putting this at issue. Plaintiffs are attempting to put it at issue. And we've consistently said that the reasons for this removal are privileged because they're based on legal advice. And it's relevant, your Honor, I think, and helpful to look at the cases on waiver because this is where we're getting. And all of the cases about waiver discuss reliance. There has to be an element that the party asserting the privilege is relying on privileged communications in part to support a claim or defense.

And that's obviously not at all what is going on here. It's just the opposite. And I would refer your Honor to the most recent decision we found is the *GLD3, LLC v. Albra* case in

this district from 2024. And I think that the language in that case is --

THE COURT: Is there a docket number for that case?

MR. SLAUGHTER: That case, *GLD3 v. Albra*, is 2024 WL 4471672. And that case quotes the Second Circuit in *In re County of Erie* from 2008 which emphasizes that a privilege waiver only happens when a client places the attorney-client relationship directly at issue. So the idea would be somehow that OpenAI is putting this quote directly at issue and asserts reliance on attorney's advice as an element for a claim or defense.

That could not be further from the truth. We have been steadfast and consistent that we are not relying in any shape or form on advice of counsel with respect to the reasons for the removal of datasets for any claim. And that's the -- and that is really -- and the real sort of dichotomy or -- I'm missing the word, but there's sort of two different issues. One is this removal of these datasets and subsequent recovery and production to the plaintiffs so they have the data.

But there's a separate question about use, right? Okay. So they were used for a period of time in training and fair use, and there will be loads of evidence on fair use. And so -- but the reasons --

THE COURT: Yes. So I guess you were about to say, but why were they removed?

MR. SLAUGHTER: The reasons for the removal are all privileged, your Honor. We've said to the plaintiffs --

THE COURT: But you're not relying on advice of counsel? They're privileged reasons because presumably they were removed -- I'm trying to understand how this isn't circular. The pirated datasets were removed, but we can't tell you why because the reason is privileged, but we're not relying on advice of counsel, but counsel told us that we should remove them.

MR. SLAUGHTER: We're not -- we are not offering any evidence, your Honor, about the removal of datasets. We're not offering them on any subject. The plaintiffs had asked us why did you remove them, and we said the reasons we've removed them are privileged. We are not advancing -- we're not relying -- we're going to go to trial. We're going to go to summary judgment. We're not advancing the reasons for removal or discussing removal.

If plaintiffs are going to want to discuss it, they have all they need to discuss it. They can say they downloaded on such and such a date, they used it to train such and such a models, they subsequently removed it, and so you can draw the conclusions you want to draw from it.

THE COURT: So then the conclusion would be that there would be -- there was some understanding or knowledge that maybe it was not really a great idea to use these datasets, so

we're going to remove them, but we're not going to tell you that that's why it was. I can't ask you to respond to that, but doesn't this go to knowledge or awareness, which is something that the plaintiffs need to prove?

MR. SLAUGHTER: I don't know if it is something that the plaintiffs need to prove, your Honor. But we are not relying on it for any claim or defense that we have to bear the burden on. So --

THE COURT: Well, but that's only half the issue, right? I mean you're not relying on a defense, yes, because maybe it sucks.

MR. SLAUGHTER: Your Honor, we're not relying -- there's --

THE COURT: So there's no reason. There's no reason why --

MR. SLAUGHTER: No.

THE COURT: -- a fact finder would then be permitted to draw an inference from the lack of evidence because the evidence is being withheld?

MR. SLAUGHTER: The fact finder can draw whatever inference that the fact finder wants to draw from the non-privileged evidence. And the non-privileged evidence is it was acquired and downloaded on such and such a date, and was used to train such and such a model for whatever period of time. It was subsequently removed and recovered. And that's

the -- what they can rely upon, and that's the non-privileged information they can rely on --

THE COURT: Wait, wait, wait. So there is an admission though that that data set was used to train GPT-3.5?

MR. SLAUGHTER: Yes.

THE COURT: Okay. So does that mean that there's an acknowledgment or an admission that infringing material was used to train GPT models that are at issue in this case?

MR. SLAUGHTER: We -- if the question is whether there was copyrighted material used to train models, that's the --

THE COURT: Copyrighted material that was obtained with or without knowledge, that it was obtained from a pirated database? I mean, I think one of the issues in this particular dispute is that at some point in time somebody figured out that these were not -- they were obtained from a not-so-great source.

MR. SLAUGHTER: That's a different question than the question that's on this motion. The question on the motion is the reasons -- the waiver based upon the reasons for removal. You should tell us what the reasons are for the removal, and we said all along that they are privileged.

The question that you're asking is a different question, which is when did you download, which we've given, and what did you know about them? They can ask all of those questions about when did you download them and what did you

know about them at the time.

THE COURT: So if a lawyer -- hypothetically, of course, because I haven't looked at the discovery and I'm not -- so hypothetically, if someone were able to take a deposition and say, okay, you downloaded these, and at what point in time did you come to find that these were acquired from LibGen? What do you know about LibGen and when and who told you? Doesn't that get you right --

MR. SLAUGHTER: I think that the -- but the question on this motion is the reasons for the removal, and so you can -- there can be questions at a deposition. You acquired this from LibGen? Yes, I did. When did you do it? How did you do it? What did you know about it?

THE COURT: Take me through every step of your knowledge.

MR. SLAUGHTER: Yeah. And I think when they get to the question about was it removed? Yes. Why was it removed? Those answers are privileged, and we've always said they're privileged. And we've said that there are no non-privileged bases for the reasons for the removal. So this is just this point at the very end about why was -- were these datasets removed, and then obviously they were subsequently recovered and produced.

But the question on this motion is why were they removed, and that question is entirely privileged. The answer

to that question is entirely privileged. We've been consistent about that all along. And we've explained to plaintiffs that there are no non-privileged reasons for the removal.

THE COURT: All right. I see Mr. Smyser shaking his head. Come on up and talk to me.

MR. SMYSER: Your Honor, there's a key point here that I think is crucial to understand. Mr. Slaughter repeatedly asserted something that is not true. They have not recovered the datasets used to train GPT-3.5. They have also not recovered the original downloads from LibGen. Those downloads are in this case. They are part of how the datasets were created. We cannot determine what works were included in them. They are, as Judge Alsup recognized in the *Anthropic* case, themselves an independent act of infringement. We also cannot determine because of their deletion the contents of the GPT-3.5 training datasets because they no longer exist because they were removed.

Secondly, he repeatedly stated that OpenAI has been consistent that these reasons were all privileged. Search your memory. You were at the May hearing. You heard their counsel say that all the reasons were not privileged. They've changed their position on this multiple times because they originally put in a non-privileged reason into the record as part of an argument to this Court. That is sufficient on its own to put it at issue.

But even if you disagree with that, and then the shifting positions which have caused us massive prejudice over the past 15 months, even if you disagree with that, there is no question that the reasons they deleted these datasets and removed them — again, the only ones they have ever done so — goes directly to OpenAI's state of mind, which is what your Honor was touching on.

This is the *Arista Records* case again cited in our papers. You did not need to be advancing a specific affirmative defense of advice of counsel in order for the state of mind to be at issue. They are arguing that they had an innocent, good-faith belief in the legality of their actions in order to contradict a finding of willfulness, which goes directly to a statutory damages enhancement.

They'll be arguing that to the jury. We need the materials in order to test that. It's a classic sword and shield issue. And the unfairness that plaintiffs have frankly already suffered at the hands of OpenAI's conduct is sufficient under the implied waiver doctrine to find a waiver here.

A few more minor points. I believe Mr. Slaughter said that they've allowed us to ask deposition questions to all of these issues. At the most recent deposition they prevented the witness from answering a yes/no question on whether or not non-privileged reasons for the existence of books one and books two being deleted existed at all.

The grounds for that instruction were entirely improper. They were that, witness, I instruct you not to answer because plaintiffs will use that to argue waiver. That's not allowing the witness to answer the question, and it's not a proper privilege instruction. Because at the time of the deposition, their position was not that there are no non-privileged reasons. Their position was we won't advance any non-privileged reasons. That was I think the fourth of their five metamorphoses over the past year.

I will also note that the *GLD v. Albra* case is much weaker than they're stating it. Again, I invite your Honor to review *Bilzerian*, *Chipotle*, *Arista*, and *Grund* from our papers, which clearly indicate that there is no requirement that the reliance be expressed or particularly invoked. The reason for that is that that would be express waiver. We're talking about implied waiver where the content of the arguments that the party is advancing necessarily implicate a privileged matter, which would be unfair to allow the other party not examine. That's where we're at here. Couldn't be clearer.

Happy to answer any questions.

THE COURT: All right. This is another thing I'm not going to rule from the bench. So I think we're winding up here. All right. Mr. Slaughter?

MR. SLAUGHTER: I was going to obviously allow you to finish, your Honor, if you had anything more to say. But there

were things Mr. Smyser said that I feel deserve a response, but if you --

THE COURT: You can respond, and then I'm going to finish up the conference.

MR. SLAUGHTER: Thank you so much. Just with respect to the *GLD3*, your Honor, it is directly a case about waiver by implication and it talks about reliance, as I just discussed.

Mr. Smyser discussed all the datasets that they say they don't have. A, that's not true. The datasets have been produced to them. B, of course 2022, at the time the datasets were removed, they were removed a year before any litigation had been threatened about training, so there was no duty at that time to retain. And I just want to --

THE COURT: Wait. But you just said the datasets were produced.

MR. SLAUGHTER: Yes, we recovered them.

THE COURT: And you're also arguing there was no duty to retain the datasets that were removed?

MR. SLAUGHTER: Correct.

THE COURT: What datasets have been produced?

MR. SLAUGHTER: We recovered the books one and books two datasets after this litigation was filed, and we searched for them and found them and produced them.

THE COURT: Okay.

MR. SLAUGHTER: Okay. And then your Honor, just with

respect to this point that they have -- and this is in our brief. We -- if your Honor's interested, there was a letter that was written, meet and confer letter and one letter that was written to the Court, and Mr. Smyser characterizes that letter as the purpose of the letter was to describe the reasons for the removal of those datasets.

I encourage your Honor to look at that letter. That's not what the purpose of that letter was. The intent of the language in those letters could have been more precise than they were, but it was clear that they were intended to disclose that the data sets had been removed. And at the time they were prior to any litigation being filed, and that OpenAI was actively investigating whether they could be located. And that's what those letter were about. They were not some letters that were intended to describe all the reasons for the removal.

Your Honor, Mr. Smyser touched on the crime-fraud -- this isn't crime-fraud. This is fair use. As Judge Chhabria found in the *Kadrey* case where he found that it was fair use to train on books that had been downloaded from LibGen. So don't want to not address the crime-fraud issue. If your Honor's interested, that's obviously addressed in our papers.

THE COURT: I don't need to hear it again.

MR. SMYSER: Your Honor, it's not true. It is not true that they have produced the books one and books two

datasets. I want to be very clear about this. There was a version of books one and books two used to train GPT-2. They located a copy of that and have produced it. There was other versions of the books one and books two datasets used to train other models including GPT-3.5. They have not found or produced that to us.

They have also not found or produced the original downloads from LibGen. Those have all been deleted. We don't have those. We cannot investigate a claim of noninfringement as to those datasets, for example. So it is simply not true to say they have covered and produced everything. That is not true. I had to make that clear because that's; the record of the case so far, and it's crucially important to understanding what's at issue here.

THE COURT: All right. August 22 you're going to -- plaintiffs are going to provide me a timeline, your timeline on the flip-flopping of the positions, your alleged flip-flopping of the positions. You're going to cite to ECF numbers if they're already here. In other words, I don't need to see them all produced again. Just give me the ECF numbers and I can look at them. All right? And then August 29, OpenAI can respond to that. And then after that, I will see if I need additional briefing or I have additional questions. Okay?

MR. SLAUGHTER: Thank you, your Honor.

MR. SMYSER: Thank you, your Honor. And the August 22

filing, to be clear, is just the timeline, correct?

THE COURT: Timeline with citations to ECF filings or if they haven't been filed yet --

MR. SMYSER: As exhibits. Understood.

THE COURT: As exhibits. And response from OpenAI August 29. Do not file more on this issue until I say so or ask for it. I am taking this very seriously. It's something that's quite concerning. And if I have other questions, I will ask them. All right. Anything else?

MR. SLAUGHTER: Your Honor, nothing on that matter. I did want to raise a housekeeping matter, which perhaps this is where you were headed, but at the risk of another month of discovery reports, it does seem to me we should schedule our next conference together. I don't know if you were thinking about that.

THE COURT: I have dates for those. All right. Everybody ready? Tuesday, September 23; Wednesday October 29; and, Thursday, December 4.

MS. BROOK: Okay. You're not going to like me. If it's just me, then by golly-gee, go forward. But Tuesday, September 23 is Rosh Hashanah, which I do imagine might impact more than me.

THE COURT: I should have been aware of that. I'm sorry.

MR. SLAUGHTER: Can we do the 30th, your Honor?

THE COURT: September 30. Does that get us past Rosh Hashanah?

MR. SLAUGHTER: We will all be atoning, your Honor. That's in between Rosh Hashanah and Yom Kippur.

THE COURT: When does Yom Kippur start?

MR. SLAUGHTER: Starts the evening of the Wednesday of the next day.

MS. BROOK: We could even do later in the week of the 22nd, your Honor, so folks that observe both days will be unable to travel until the evening of the 23rd. But if we did it on the 26th, that Friday, that would not be conflicted by the holidays.

THE COURT: But the 25th also would not be okay because the travel restrictions isn't lifted into the evening?

MS. BROOK: That's not ideal, no.

THE COURT: What if we did the 25th in the afternoon? Does that help?

MS. BROOK: With that one it's probably just me coming from Los Angeles, and so I will work that out if need be.

THE COURT: So maybe September 25 in the afternoon. Let me just take a moment also.

All right. The reason why I pushed for the 25th in the afternoon rather than putting it over till earlier the next week is because if we need to have another one of our settlement conferences or conversations, I'd like to try to

keep Fridays open for that. And that hopefully means that people wouldn't have to stay over the weekend if they're coming in for the conference.

So let's for now put it September 25 in the afternoon. Let's start at 2:30. Is there any issue with the other dates?

MS. BROOK: Nothing that I see of that kind, your Honor.

THE COURT: I mean, they're far enough out that if there are holidays that didn't appear on my calendar or on my radar at the time we set these dates, we can reassess. But for now, September 25 in the afternoon. We're going to start at 2:30. And then October 29, December 4. Okay. What's next?

MR. SLAUGHTER: We'll all go get our flights.

THE COURT: Oh, you're not coming back on Friday?

MR. SLAUGHTER: I will not be here on Friday, your Honor.

THE COURT: I guess there's only a subset of you.

Anything else we need to do at this time then?

MS. BROOK: Nothing else from plaintiffs, your Honor. Thank you and the court for staying so late. We appreciate it.

THE COURT: Thank you all for your stamina. Anything from defendants?

MS. YBARRA: No, your Honor.

MS. HURST: No, your Honor. Thank you.

THE COURT: Thank you very much. We are adjourned.