Pages 1 - 45

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

Before The Honorable Vince Chhabria, Judge

RICHARD KADREY, et al.,	)	
	)	
Plaintiffs,	)	
	)	
VS.	)	**NO. 23-CV-03417-VC**
	)	
META PLATFORMS, INC.,	)	
	)	
Defendant.	)	
	)	

San Francisco, California
Monday, November 10, 2025

TRANSCRIPT OF PROCEEDINGS

APPEARANCES:

For Plaintiffs:

BOIES SCHILLER FLEXNER LLP
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
BY: MAXWELL V. PRITT, ATTORNEY AT LAW
MARGAUX POUEYMIROU, ATTORNEY AT LAW

LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
Embarcadero Center West
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
BY: DANIEL M. HUTCHINSON, ATTORNEY AT LAW

For Defendant:

COOLEY LLP
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111-4004
BY: KATHLEEN R. HARTNETT, ATTORNEY AT LAW

(APPEARANCES CONTINUED ON THE NEXT PAGE.)

REPORTED REMOTELY BY: Kendra A. Steppler, RPR, CRR
Official United States Reporter

APPEARANCES: (Continued)

COOLEY LLP
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004
BY: PHILLIP E. MORTON, ATTORNEY AT LAW

COOLEY LLP
1333 2nd Street, Suite 400
Santa Monica, CA 90401
BY: BOBBY A. GHAJAR, ATTORNEY AT LAW

COOLEY LLP
3175 Hanover Street.
Palo Alto, CA 94304
BY: ELIZABETH L. STAMESHKIN, ATTORNEY AT LAW

PAUL, WEISS, RIFKIND, WHARTON & GARRISON LLP
535 Mission Street
San Francisco, CA 94105
BY: ANNA M. STAPLETON, ATTORNEY AT LAW

Monday - November 10, 2025 **2:04 p.m.**

P R O C E E D I N G S

---o0o---

THE COURTROOM DEPUTY: All rise. Court is now in session, the Honorable Vince Chhabria presiding.

Please be seated.

Now calling Civil Case 23-3417, Kadrey, et al. v. Meta Platforms, Inc.

Will counsel please come forward and state your appearances for the record, starting with the plaintiff.

MR. PRITT: Good afternoon, Your Honor. Maxwell Pritt, Boies Schiller Flexner, for the plaintiffs.

THE COURT: Hi. Feel free to introduce your team or everybody can introduce themselves. Whatever you prefer.

MR. HUTCHINSON: Good afternoon, Your Honor. Daniel Hutchinson of Lieff Cabraser Heimann & Bernstein.

THE COURT: Hi.

MS. POUEYMIROU: Good afternoon. Margaux Poueymirou, Boies Schiller Flexner.

THE COURT: Hi.

MS. HARTNETT: Good afternoon, Your Honor. Kathleen Hartnett from Cooley, for defendant Meta Platforms. And also with me from Cooley are Mr. Ghajar -- Bobby Ghajar -- Phil Morton, Liz Stameshkin. And then we have Anna Stapleton from Paul Weiss. We also have our client here.

THE COURT: Hi.

So I read your briefs last week. I read them again this morning. I'm not embarrassed to say that I have no idea what you're talking about. It's very hard to follow, because you all are living this case day in and day out. And, you know, say a lot of things that -- you know -- use a lot of terms and discuss a lot of subjects that you are intimately familiar with, but that I'm not.

So I thought maybe the most efficient way to kind of get -- cut through it all is just to have a -- have you in here and talk it through and see if we can figure it out.

So -- I don't know -- maybe let's start with you, and you can sort of tell me what you think you need and why. And I have read your brief twice and their brief twice. So I'm not totally unfamiliar with it. But I do have difficulty wrapping my brain around it. So...

MR. PRITT: Thank you, Your Honor. I appreciate it. I'm, of course, happy to answer any questions as they come up.

I may start with focusing on the -- what I'll call the "torrenting" or "peer-to-peer file sharing-related bucket." There's that bucket of information we're pursuing. And then there's a separate bucket of what I'll call the "ML Hub/AI dataset collection."

THE COURT: Okay.

MR. PRITT: On the torrenting issue, nearly everything

that has --

THE COURT: Could I --

MR. PRITT: Yes, please.

THE COURT: I'm sorry to disrupt your presentation, but, I guess, the -- I would like to hear a little more about the ML Hub bucket; right? Because if it's not related to torrenting, what is it related to that we still need to do in this case?

MR. PRITT: Well, it's certainly related to torrenting in the sense that this is the document that, for the first time, we believe shows that torrenting was initially blocked at Meta from what we have been referring to as "shadow libraries." And then, at some point, it was unblocked, and the torrenting occurred. And within that context, this is where there's redactions for attorney-client privileged advice.

Previously, one of the Meta's 30(b)(6) deponents had testified that torrenting was not part of the approval or disapproval process, one way or the other.

THE COURT: So you're interpreting -- can you tell me a little -- when you say that there's a document that says it was blocked, what does it -- can you provide a -- give me a little better description of that?

MR. PRITT: Yes. So this is Exhibit J. And what the -- what we now know -- because we had never seen these printouts prior to late August of this year -- this contains

toolkits -- I'm not sure how Meta describes it -- for -- each dataset included the ones from pirate marketplaces like Anna's Archive.

So you'll see, at the top, it says "Anna's Archive." They then produced, after we asked them -- without receiving response about other such datasets -- they had produced, a month later, ones for LibGen --

THE COURT: Wait, sorry.

MR. PRITT: Yeah.

THE COURT: So this document that you've just pointed me to --

MR. PRITT: Yep.

THE COURT: -- Exhibit J -- when did you receive this?

MR. PRITT: I believe we received this August 28th, 2025. It was --

THE COURT: Okay.

MR. PRITT: -- the first time we had seen a document such as this.

THE COURT: Okay. And talk me through it.

MR. PRITT: Well, this is the document that originally they had redacted what's on pages 6 and 7 of this document. There are updates by an employee named David McAneny. And then they clawed back one part that had not been redacted. And then we fought in front of Judge Hixson, and he ordered the production.

And what they had clawed back is -- if you're there, it's a June 28th, 2024, date on page 6 of 8. It's Bates Number 238413. One, two, three, four lines down, it says: Each dataset is hundreds -- and it's quoting something -- we don't know where that comes from -- each dataset is hundreds of gigabytes or terabytes of data unable to download due to extreme size of datasets. In addition, full downloadable datasets are accessible via torrent, comma, which is blocked.

THE COURT: "Which is blocked"?

MR. PRITT: "Which is blocked."

THE COURT: Okay.

MR. PRITT: And later on, on the next page, it says --

THE COURT: Hold on a second. I have -- I'm looking in my folder, and I'm seeing -- right now, I'm -- I was looking at Exhibit J.

MR. PRITT: Yes.

THE COURT: But it was -- it's the redacted version. So I -- whatever it is you were reading to me, I didn't -- I don't have that. So hold on.

MR. PRITT: There's a lot of redaction, but the unredacted part is at the bottom of -- where it's Bates labeled 238413.

THE COURT: Okay.

MR. PRITT: And so it's -- you'll see there's an update from someone named David McAneny.

THE COURT: Okay.

MR. PRITT: Four lines down, where it says: "Updated the 'Pii Present Text' to."

THE COURT: "Accessible via torrent, which is blocked."

MR. PRITT: Correct.

THE COURT: Okay.

MR. PRITT: A month later, Meta produces other screenshots for Library Genesis and then for --

THE COURT: And you're saying you received this document that we're looking at now -- this Exhibit -- what was it? -- J.

MR. PRITT: Exhibit J, August 28th, I believe --

THE COURT: Okay.

MR. PRITT: -- of this year. And then we had to fight the clawback. And then we got this document with the unredacted text that I just read in early October.

THE COURT: Okay.

MR. PRITT: At that same time -- around that same time -- Meta produces similar documents for Library Genesis and then another dataset that we don't know much about called Library Genesis-Anna Archive. And then, just on Friday, they produced an additional dataset for Internet Archive-Anna Archive. Three of those have this exact same language about torrenting being blocked. It's the same update.

What Meta says in its opposition is: Well, it's ambiguous.

THE COURT: It was ambiguous.

MR. PRITT: Yes. But certainly ambiguous enough to allow some additional discovery to determine what -- the reference to torrenting being blocked and then, at some point, presumably being unblocked, since they did, in fact, torrent this.

THE COURT: Okay.

MR. PRITT: The other issue related to torrenting is just that torrenting certainly also means downloading. And these dataset files show -- including the MD5 hashes of all the books downloaded from these pirated databases. They haven't produced those files yet. We've asked for follow-up, and Meta's response is that discovery is closed.

So the ML Hub and the AIDC are related to torrenting, because they are about the content that is downloaded, and then they specifically reference torrenting.

Now, as to the actual torrenting-related files --

THE COURT: Could I ask -- could we --

MR. PRITT: Yes.

THE COURT: -- stay on the ML Hub a little more?

MR. PRITT: Yes.

THE COURT: Do you know what the ML Hub is overall? Like, what is it?

MR. PRITT: We don't really know.

THE COURT: And what is contained in it?

MR. PRITT: I asked that when we first got this. And I asked if it was a noncustodial data source. And, eventually, Meta's counsel confirmed that it was what we, plaintiffs, would understand is a noncustodial data source. It was not one that was disclosed by Meta when they disclosed their noncustodial data sources that they searched. It came up obliquely in some depositions in late November, well after the last day to serve any discovery requests in this case. We had never seen anything like this document that you just looked at.

As to what the -- the AIDC, which is AI Dataset Collection, is just a part of the ML Hub. We don't know what else is in the ML Hub, although there is a document that was produced, I believe on the last day of discovery, that said it is the unifying source of all dataset acquisition -- or all data acquisition -- and it looks like experimentation on everything that is run on the data that Meta obtains, at least for Llama 4 and onward.

THE COURT: Okay. Now, speaking of Llama 4 -- that was going to be my next question -- Meta says something in its brief about how this -- all of this stuff that you're asking for now relates to how Llama 4 was trained. And that Llama 4 -- whatever you did to train Llama 4 is not part of this case. I'm sort of -- that -- Meta didn't exactly say that, but

I was reading between the lines of their brief; right?

Apparently there was some order by Judge Hixson that they didn't have to disclose all the Llama 4 stuff. I interpreted that discussion to mean that this -- that, at some point, everybody came to an understanding that the way Meta trained Llama 4 is not the subject of this lawsuit. That if you want to go after them for what they did in connection with Llama 4, that's a different lawsuit. And, if so, that would make sense to me; right?

It's like -- you know -- it's like these patent cases, where if somebody sues Apple for, you know, patent infringement and -- you know, for some -- for the iPhone, and we keep getting new versions of the iPhone, and, at some point, you just have to stop and say we're going to have a trial about, you know, the iPhone 14 and lower, but we're not going to have one about the iPhone 16. And that's going to be the next case. If you want to bring another case about the iPhone 16, you can.

So I was sort of assuming that that's what happened with respect to Llama 4. Was I reading too much into those statements in the briefs?

MR. PRITT: Thank you, Your Honor. Yes. This was the exact issue we did fight in discovery in front of Judge Hixson. And he held that both Llama 4 and Llama 5 were within the scope of the case.

THE COURT: Okay.

MR. PRITT: In fact, all of the datasets in the -- the data at issue from these pirated datasets were obtained by Meta and torrented before the close of fact discovery.

If you look at Docket 279, which is what Meta cites -- it's Judge Hixson's order -- it says: The Court agrees with plaintiffs that Llama 4 is relevant to this case, notwithstanding that it is still under development.

And when Meta cites this docket in the next sentence, they say, in their opposition, that Judge Hixson held they didn't have to do any searches for Llama 4. That's not what he held. What he held was, based on Meta's representation that its search terms would encompass all documents about Llama 4, because Meta did not limit its searches to any particular version of Llama, Meta just wouldn't have to do anything else.

So -- and, in fact, all of the documents we have gotten -- including just 48 hours ago, an additional 55 terabytes of copyrighted content torrented from LibGen Non-Fiction, which contains over 50 asserted works in this case, as well as some derivatives of those works -- really, all of that is responsive to RFP 7 through 9 in the very first set of discovery issued in this case, which asked for all documents and communications to, from, or with Library Genesis, Z-Library, and another pirated dataset.

They're certainly all relevant to the baseball arbitration RFP that was issued after your Court's order in March. And

they're, almost all of them, noncustodial data sources that Meta repeatedly represented. If you look at Appendix C on our motion, that it had searched all relevant noncustodial data sources and produced documents.

And not responding -- notwithstanding the responsiveness of these documents, we've been asking about torrented data since we first learned about Meta's torrenting in the second week of November, when we took a 30(b)(6) deposition, and, for example, with the 2022 downloads of Library Genesis by an employee named Mr. Lample.

THE COURT: Let me interrupt you again just to seek some clarification.

Okay. So you started off by putting it in two buckets, torrenting and -- the torrenting bucket and the ML Hub bucket. And I sort of interrupted you and I asked you to start talking about the ML Hub bucket. Can you just step back for a second and tell me -- so you want to -- what discovery are you asking for with respect to the ML Hub?

MR. PRITT: In our motion, we are asking for them to search the ML Hub. We provided some search terms that we believe are narrowed to torrenting and shadow libraries.

THE COURT: Okay.

MR. PRITT: It's Appendix C. And they largely overlap with the search terms that we had proposed in our baseball arbitration RFP. And so we're asking that Meta run that

against the ML Hub, including the AIDC, and then some 30(b)(6) questions about the ML Hub.

Because the AIDC, for example, in the ML Hub, states what the possible uses are for different datasets, which, of course, is very relevant to the Court's inquiry. It says training, experimentation, and then, just from Friday, a new use that is titled Prod/OS Training Model. I don't know what that means. Those are the types of questions --

THE COURT: I guess what I'm confused about is what you have remaining is this distribution claim that is -- that is about torrenting; right? That is about -- it's about whether -- it's not -- it's not even about whether you downloaded stuff from the shadow libraries. It's about whether -- not you -- Meta -- it's not about whether Meta downloaded stuff from the shadow libraries. It's about whether Meta uploaded stuff as it was downloading stuff -- or in the immediate wake of downloading stuff from the shadow libraries.

And it seems like most of what you're talking about, if not all of what you're talking about with respect to the ML Hub, doesn't relate to that.

MR. PRITT: Right. Thank you, Your Honor.

So I would put that into -- in two different things. First, torrenting -- when we talk about that, that involves three different things. It involves making copies while downloading, necessarily involves, based on the settings Meta

used, uploading. So it's copying, making additional copies while uploading and distributing, and then also making the filings available. So there's three separate methods of analysis under what we're calling "torrenting."

So certainly downloading, because they did it by torrent and also uploaded at the same time, is still relevant to this case.

THE COURT: Is that established? I mean, I remember --

MR. PRITT: Yes.

THE COURT: -- that there was -- when we were at the last summary judgment motion, there was still some question whether Meta was uploading stuff as it was downloading it.

MR. PRITT: There is no question, including based on Meta's own torrenting expert, that Meta was uploading --

THE COURT: Okay.

MR. PRITT: -- while downloading --

THE COURT: Okay.

MR. PRITT: -- mass amounts of copyrighted works. I think their expert said 40 terabytes --

THE COURT: So anything Meta was torrenting during whatever time frame we're talking about -- training Llama 4 -- whatever -- there was uploading happening at the same time as downloading?

MR. PRITT: Correct. The dispute and the issue will

be really an evidentiary one, because the way BitTorrent works, it doesn't create -- unless you're actually joining the swarm. Like some, you know, universal music group or some others will actually go in and join swarms with their works to try and see if you're downloading their works. Unless you're taking that, the way BitTorrent works is you're never going to have the data as to what precisely was uploaded, unless you're actually writing code to do so, which no one does, because you use BitTorrent to remain anonymous and not have that trail.

So, instead, the question is going to be: What percentage of each work was downloaded? Because when you're torrenting from these systems, you're taking many copies of an asserted work, for example. You might have a dozen copies of Sarah Silverman's work at issue in this case.

And then it will be a statistical analysis as to what's the likelihood that that work was part of the 40 terabytes plus that Meta uploaded while it was torrenting about 130 terabytes. And when I say those numbers, those are the numbers that Meta's expert used. We now know there was a substantial amount of additional torrenting based on the productions since late August.

There was something else you'd asked.

THE COURT: So you're just saying that any -- any evidence of downloading from a shadow library is going to be evidence of distribution?

MR. PRITT: Based on the settings that Meta used, certainly it would be, at a minimum, circumstantial evidence. And because we don't have a clear record, based on how BitTorrent technology works, of exactly what was uploaded, we have to understand, in order to piece different pieces together, what exactly what downloaded when it was downloaded.

We have to go find usage logs and any other types of download logs to tie together how much is being downloaded, how much is -- what amount of traffic is going out through Meta's servers -- or Amazon's servers that Meta is using -- in order to try to piece together what is really a puzzle of exactly what is being uploaded versus what was being downloaded.

So it's very --

THE COURT: Okay.

MR. PRITT: -- different, for example, than if Meta went piecemeal on the Library Genesis and clicked single files to download directly.

And the way I read your summary judgment order -- could be mistakenly -- was that perhaps if you're doing that, and that the ultimate use is AI training, that could be fair use. But you left open the question: Well, what if you're doing it by torrent? Where you're going in and downloading mass quantities, and, at the same time you're doing that, uploading and distributing pieces. And, again, when you're uploading, you're both making additional copies, making the works

available --

THE COURT: To the extent that you're suggesting that that could change the fair use analysis, I think the answer is no. You're done with fair use; right? This is -- what you have left is a distribution claim. And you -- you know -- you're making noise about sort of revisiting the fair use ruling. I think what I made very clear in the fair use ruling is you can't win without making a presentation on that fourth factor. And we're not talking about the fourth factor anymore. So --

MR. PRITT: Correct. I'm not asking about --

THE COURT: In your brief, you've alluded to how this stuff is relevant to what I already ruled on a couple of times. And I just want to make very clear that it -- it might be. But we're not revisiting -- like, you've had your chance on fair use. And you lost. And so now we're looking at distribution.

MR. PRITT: I understand, Your Honor. The only thing I would say to that is, if Meta failed to disclose highly relevant non-custodial sources to that question, we would submit that Meta should provide that information now to provide us the opportunity to determine whether or not it is relevant to our analysis. If it should have been produced before your summary judgment order, then it should have been produced. And we should have the opportunity to look at that evidence. But --

THE COURT: Well, but the reason -- the primary reason we may have somewhat of an incomplete record for the fair use analysis is because of your side and your side's failure to litigate this case properly. And I don't want to get into that with you again. But -- and whatever you find on torrenting, it doesn't -- it's not relevant to the fourth factor. And the fourth factor is why -- is the main reason you lost on --

MR. PRITT: Correct.

THE COURT: -- fair use.

MR. PRITT: And Judge Hixson denied our request for evidence relevant to the fourth factor. But that's neither here nor there at this point. We're focused on the torrenting evidence and the remaining claims in this case after your summary judgment order. And we do think the ML Hub and AIDC are relevant to that, because they specifically talk about the torrented data from the libraries that are at issue still with respect to the uploading claim.

And I just want to make clear, there are three general uploading periods -- downloading and uploading periods: 2022, which was by an employee named Guillaume Lample. This is October 2022 to January 2023. This is Library Genesis -- what are called the Fiction and Non-Fiction or Scitech databases.

We originally had been seeking -- this was the original torrenting, I believe, that we learned about in mid-November. We had been seeking records. We were repeatedly told that

Meta's 30(b)(6) depositions -- that they investigated and could find nothing. And then all of a sudden, in April, they produced 3.9 terabytes of logs showing torrented data from Library Genesis. Based on the records and what we understand Mr. Lample to have torrented at that time, they're -- it's missing about 60 terabytes of torrented data.

THE COURT: Could I just interrupt you for one second and say --

MR. PRITT: Yes.

THE COURT: -- you -- you said that Meta previously told you, "We investigated and we found nothing."

MR. PRITT: Yep.

THE COURT: Could you show me where that is?

MR. PRITT: Sure.

If you look at Appendix B to our motion, the second page, you'll see part of the chart refers to, again, Lample 2022 to 2023 Torrenting.

THE COURT: Hold on. Let me make sure I'm on the same page, literally, as you. So page 2 of 4?

MR. PRITT: Correct. At the top, there should be a second row: Guillaume Lample 2022 to 2023 Torrenting.

THE COURT: Which appendix is this?

MR. PRITT: This is Appendix B.

THE COURT: Okay.

MR. PRITT: You'll see that my colleague --

THE COURT: What is this document?

MR. PRITT: This is an appendix showing representations about -- prior representations about investigations and productions of torrenting data or noncustodial sources, and then what we have recently been told after they have, in fact, discovered --

THE COURT: Okay.

MR. PRITT: -- torrented data.

And so, in fact, for the 2022 to 2023 torrenting, it was first raised by my colleague here today in November -- we don't have that cite, but she could get it -- same 30(b)(6) witness -- and she asked if he would go investigate and come back to the next deposition a week later. He said he would. He didn't find anything. We asked again, when he was deposed a third time, on March 3rd, and they said they hadn't tracked anything down.

The March 6 discussion is from Ms. Frederickson-Cross' -- Meta's torrenting expert -- first deposition, where she also says she hadn't found any records. And then a month later, in early April, they produced some samples of logs from Mr. Lample and then produced all of the logs two days before -- sorry -- all of the 3.9 terabytes of logs two days before the summary judgment hearing.

But, again, the records indicate to us that approximately 70 terabytes were downloaded in full. So we have no idea where

Meta ultimately found these 3.9 terabytes of logs. And we have additional directories and locations that we have proposed to Meta to search to find where we believe and our expert believes the remaining logs exist.

THE COURT: Okay.

MR. PRITT: And I also want to point -- while I'm finding the exhibit number, there's a document that's called King Exhibit 1. I'll get you the exhibit to our administrative motion. That during Meta's 30(b)(6) witness testimony in early April 2024 -- this was in connection with the baseball arbitration discovery -- produced a document that essentially purports to detail Meta's torrenting for 2022, 2023, and 2024, which we now know, based on the productions from the last two months, is incomplete. But this also talks about the 2022 downloading we just discussed.

2023, we have torrenting from February 2023 to June 2023. Also more LibGen Non-Fiction and what's called LibGen Scimag. We have no logs whatsoever for that torrenting. We have followed up with Meta suggesting areas based on our expert's analysis of where such logs might exist. Meta, again, has said discovery is closed; we're not going to respond.

And then, in 2024, we have April 2024 to September 2024 torrenting. This is mostly through Anna's Archive. This contains Library Genesis Fiction, which Meta and its expert previously said was not torrented, both in, I believe, 30(b)(6)

testimony and certainly the expert's report.

We have Library Genesis Non-Fiction, which previously was represented, again, both in the document I showed you, plus the expert testimony and report, that only 10.3 terabytes was downloaded from Non-Fiction. But now we know, in fact, it was over 60 terabytes, just based on Friday's production. And then there was also torrenting from Z-Library, Internet Archive, and a database called Duxiu.

But, again, there, we still do not have all of the evidence, because, in April of 2024, there's an unexplained 20-to-30-terabyte download that we have not yet gotten to the bottom of, although we thank Meta for agreeing to investigate that issue.

The King exhibit I showed you, Your Honor, was Exhibit T to our administrative motion.

And, surprisingly, this nonfiction data that was just produced -- this 55 terabytes -- comes from the same storage location in what's called S3. S3 is a noncustodial data source that Meta, in fact, did disclose in September '24 as having searched and produced relevant data from.

These nonfiction files come from the exact same repository as the other files that Meta produced that it had torrented from Anna's Archive. So it leaves us wondering why these files were not also produced and what else might be located in both those common directories that are on S3 and then also the

personal directories of the engineers who are involved in torrenting, which, again, we've asked Meta for, but Meta has not produced.

And then every time we get new productions, like we did this Friday -- they produced command history that we've been asking for, for two months -- the command history shows the code that's being executed for the torrenting. That then shows there's an additional directory in that command history we just got that shows additional torrenting for which we have no logs.

Now, this is something -- you know -- we've proposed what we believe is narrowed discovery, including 30(b)(6), to correct and clean up the record, because we now have 30(b)(6) testimony and expert testimony from their expert, including the report, that is completely inconsistent, if not -- doesn't take into account all of this new evidence they produced.

So what we're trying to do with this discovery that we've presented is do something other than a reopening of discovery, like we had proposed in January of this year, when Your Honor accepted Meta's proposal instead. We're trying to do something targeted, based on what we have discussed with our expert, to, as best as possible, get the full record on Meta's torrenting activities.

You know, an alternative would be to allow our expert to go in and examine the S3 files and the backups of these three engineers that they have. All our expert would have to do is

easily go into the S3 repository -- in what's called "fair-spark" -- and search it for .torrent.

We'd search that across the common directories and locate everything, including everything that's been produced in the last two months and everything that's missing, and then also search the personal directories in S3 that are associated with those engineers who torrented. And then look at the clones or backups that we know Meta now has, which we only learned recently from Mr. Lample's work computer. He left in 2023. And then also Mr. Bashlykov and Ms. Wang's computers, because they're the ones that performed the majority of the torrenting at issue.

And then if Your Honor would allow our expert to do that, the only other thing that Meta would have to do is produce usage logs for the torrenting periods. Just those two things.

We assume Meta would object to that, and perhaps Your Honor wouldn't go for that, so we tried to, instead, provide the narrowed RFAs, which are all RFAs that our expert prepared, that only Ms. Wang could answer. Because they're all about the torrenting that she performed in 2024. We've asked for her limited deposition in conjunction with those RFAs and then a 30(b)(6) that has discrete topics, mostly geared towards the torrenting, but we did include some -- to ask questions about the ML Hub and AIDC --

THE COURT: If you're going to take her deposition,

why do you need the RFAs also?

MR. PRITT: Because --

THE COURT: That may be a dumb question. And I apologize if it is.

MR. PRITT: No, it's a good question.

Because, right now, the record is incomplete based on the 30(b)(6) testimony and the expert testimony as to exactly what was done with respect to torrenting. What this does is allows us to take a seven-hour deposition, cut it in half, and ensure that she has the time that she needs to investigate, rather than coming to the deposition and saying, "I don't remember. I don't know."

And then that would enable Your Honor to cut through a lot of the issues when it comes to torrenting and focus really on is the making available -- right -- one that the Court's going to recognize? Is circumstantial evidence sufficient to prove distribution when it comes to use of peer-to-peer file sharing technology? All we're trying to do is cut down so that you can focus on those issues when it comes to summary judgment.

THE COURT: Okay. Thank you.

MR. PRITT: Thank you, Your Honor.

THE COURT: Ms. Hartnett, do you have anything you'd like to say?

MS. HARTNETT: Where to begin. No, I -- sorry that our brief wasn't clear. I think --

THE COURT: It was not -- I don't think it's --

MS. HARTNETT: I thought it made sense, so it means I've definitely been fully sucked into the details here. But, no, I appreciate the point, because it is hard to know what matters. And there was a lot I do want to respond to, but I also just want to help focus the Court on what we think has happened.

And I have -- like I said, I'm working closely with our folks that are doing the discovery on a day-to-day basis. And a lot of what we just heard is not fully accurate. But I'm just trying to think of the best way to help focus you.

And one of the things that just comes to mind, I think, is he just mentioned that the expert wanted to look at Bashlykov and Ms. Wang's computer. But the actual request in the motion was for Mr. Lample's. So the -- what we were just hearing actually was an evolving set of requests from even what was in the motion.

I think where -- if I may just start with the AIDC/ML Hub issue, which I -- that really goes to the Rule 72 motion before Your Honor, where Judge Hixson -- I think -- it's not fully briefed up yet, but that's one where we're briefing. There's a document that had privilege redactions. Judge Hixson lifted one set at the back that was from a nonlawyer. And there's a stray reference to torrenting there.

And I think that the plaintiffs have sort of taken that to

mean that the legal advice in the front of the document somehow was about torrenting or unblocking it. And we've explained that's not actually -- that's -- they're two separate parts of the document. So we think they're overreading that document.

But, more importantly, the reason why they even have that document -- the AIDC/ML Hub page -- is because it was actually hyperlinked in a document that was responsive to their baseball -- we call it the "baseball arbitration RFP." Under our protocols, we actually agreed not to put -- hyperlinks were not treated as attachments, but really taking Your Honor's order to heart of "when in doubt, produce."

And I think you can see it, for example, in Exhibit S to the Pritt declaration, where we have one example of our emails back and forth with them. We have really tried to go out of our way to do that. So we did a go-get. We got the page from the AIDC related to the dataset. And as we tried to explain in our brief, those are Llama 4 generated tools. And so they weren't originally part of our search and collection. And we -- what Judge Hixson said -- and I think it's important for this order -- you don't have to go run new searches for Llama 4, but Llama 4 is in the case.

Moreover, it really wouldn't have -- again, the RFP from the baseball arbitration was sufficient to show what stuff you torrented, essentially. And that is what we have been working super hard to do. We've, I think, gone beyond a reasonable

search. We heard some critique, for example, that we, at one deposition, didn't yet have Mr. Lample's logs -- pardon my French -- and eventually we did get the Lample logs -- some of them.

We actually had to go through a backup of his computer systems. We went back and looked more after our witness had said: Not that we don't have them, that we haven't been able to find them yet. We went back and looked through his backups. He actually left the company before this suit was filed. He's in France. He runs a competing AI company. But we did find those logs. And they were references to terabytes of logs. Those were data. So we have found what we can find of what Mr. Lample torrented.

And I think the one other -- so just I hope -- I know I'm kind of sinking back in the weeds. It's so hard not to. But I -- but I do think that's just an example of how we actually have made not just reasonable, but, like, exhaustive efforts, to find what was torrented so they can have it and do their probabilistic analysis.

The part that I do think is fair -- and we tried to make clear of this in our motion -- Ms. Wang had a dataset that she torrented. We disclosed that before. There have been a couple of additional recent disclosures about similar work that she did.

One was sort of a rerun of one of the torrents because --

to capture non-English documents. And a couple of them were test runs earlier on of 6,000 small run for fiction. There was another one -- I think we explained this in our brief -- there was a small one about seeing if she could do it through topping off the work that another person had done.

So it's not as if there's a whole new group of people at issue. There's discrete universe of Lample, of Bashlykov, of Ms. Wang. And we've really tried to bottom that out. And then the things that they were mentioning about follow-ups, we investigated the data spike, and that's actually how we came to find the second torrenting episode.

So I'm trying to stay high-level to not get sucked in. But I actually really believe, and we genuinely have tried our best, to comply with the baseball arbitration, with the spirit of disclosing -- erring on the side of disclosing -- that the Ms. Wang deposition will answer the questions that they were not able to ask her because they didn't ask for her as a deponent in the baseball arbitration. But we, yet, understand you want to have a full record. They should be able to ask her about that.

The RFAs actually are not something that Ms. Wang will be responding to on her own. A lot of them call for complicated expert analysis. There's 72 of them. So, again, we think that's something their expert can do. It's essentially trying to shift that work to us.

And I'm not trying to -- I know you don't want to relitigate the baseball arbitration, but that literally is something they could have asked for -- the ability to serve 72 RFAs to confirm whatever downloading stuff you produced -- and they didn't. And they had a more modest proposal that won the day then.

I'm just trying to see if -- those are the main points I'd like to make. I mean, I think -- hopefully Your Honor can see through this. That the main thing we agree that has come and flowed since they had a 30(b)(6) deposition after our baseball arbitration disclosures is Ms. Wang's additional downloading episodes, which we've disclosed as soon as we were aware of them. And that's something they should be able to question us about.

But in terms of trying to essentially boil the ocean of other parts of the company, especially parts of the ocean like the whole Anna's Archives datasets that Judge Hixson had already declined to allow them, we just don't believe there's anything further there.

And, at some point, even though we are trying to err on the side of disclosure, in the spirit that you've expressed that we've taken to heart, enough is enough, and we just think it's time for the experts to crunch the numbers once they have the Wang information --

THE COURT: Well, so their motion is premised on the

idea -- I mean, one of the things that Mr. Pritt said is we're just trying to get a full record of Meta's torrenting activities.

MS. HARTNETT: Yes.

THE COURT: All right. And it seems to me that the best way to adjudicate the distribution claim is to have a full record of Meta's torrenting activities; right? And -- and then -- so if we don't have that, the question is, why don't we have it, and what more would need to be done to get it? And I think we've established -- right -- that one of the things that needs to happen is Ms. Wang's deposition needs to be taken again. You've disclosed additional torrenting activity by her.

MS. HARTNETT: Yes.

THE COURT: And her deposition -- at a minimum, her deposition needs to be taken so they can ask her questions about that torrenting activity.

Is there any reason to believe, from your standpoint, that there's any other torrenting activity that Meta engaged in, in connection with this AI project, that we don't yet have an understanding of?

MS. HARTNETT: No, Your Honor, not based on our, I would say, reasonable-plus search. We had the Lample download, and they believe -- they speculate -- that there is more than what he actually ended up torrenting. But the records are what they are. We've gone through his backups. And so -- we've

looked at the platforms -- the platform -- that he was known to have used to torrent. So, no, not for him.

You have Nikolay Bashlykov, which was not even including plaintiffs' books. It was not fiction. But, in any event, that I don't think is disputed. It wasn't part of the motion.

And then for Ms. Wang, there is the -- almost all of it had been disclosed -- you know, not all of it -- but some of it's been disclosed post any opportunity to depose a Meta witness. And so that's why we're saying Ms. Wang.

And to be clear, this is flowing from an investigation where we were going in to talk to the people that are on the GenAI team. We had to figure out what datasets went into the training, which is what the whole other part of the case was about. And we've done that. We've looked for those people. We've -- they have the work chats, the emails to try to piece together, is there someone else out there that we think have reason to believe, in the AI world, did this? And no. The answer's no.

And the -- there's one stray -- like, we're trying to not -- like a random employee -- it looks like a temporary worker or somebody may have torrented. So that was the other spike investigated and not borne out.

So what they're really asking for -- one of their requests is to go across the entire company for their AWS -- you know, Amazon instances -- to search all of that to see if anyone

anywhere might have had a data spike. And we just respectfully submit that's not -- there's no reason for that. It has to -- we are at the point where once they have the -- we concede we disclosed Wang information that we had to piece together lots of clues to figure it out. They should be able to do their probabilistic analysis based on that.

And I know it's --

THE COURT: You're not going to -- there's not going to be any argument from your side that the baseline that they're using from the probabilistic analysis is unreliable?

MS. HARTNETT: We haven't seen their new analysis yet. So they'll -- we have expert reports coming. But I don't -- I don't think we're going to --

THE COURT: I mean, to put it more bluntly --

MS. HARTNETT: Yeah.

THE COURT: -- you're not going to say that their information about how much that we torrented is incomplete or something like that.

MS. HARTNETT: I don't believe so. We're trying to give them everything we can to show what people in the AI group -- really just three people -- torrented. But let me just confirm that with my colleagues that are working with the experts.

I think -- we're not going to complain that they haven't -- I mean, they may not have enough of a probability to

actually suggest that any of the plaintiffs works were torrented. But we'll have the evidence we have, and we're giving them all of the torrenting related to Ms. Wang. And there's no one else -- I don't believe anyone else that's been identified as potentially having done so.

THE COURT: On the ML Hub issue in that document in Exhibit J that he was showing me --

MS. HARTNETT: Yes.

THE COURT: -- again, this might be a dumb question, but why does it matter -- let's just say that that document stands for the proposition that Meta leadership was blocking people from torrenting up to a certain point. Why would that matter -- how would that be relevant to the analysis -- the distribution -- analysis of the distribution claim?

MS. HARTNETT: So I think that they, in their brief, sought to say that that was a corporate imprimatur --

THE COURT: Yeah.

MS. HARTNETT: -- of Meta's torrenting. Because there was this reference in that document to block torrenting. That's the only -- right -- so I think they want to argue that that is a corporate imprimatur. And I would assume they're going to say that it makes it worse.

THE COURT: They decided to unblock it at some point.

MS. HARTNETT: Right, which is all speculation. And as I represented --

THE COURT: But how does that -- how is that relevant to the distribution claim?

MS. HARTNETT: It doesn't appear relevant. I mean, especially when you look at what the discovery was that they asked for from the baseball arbitration, which was to give them documents sufficient to show all the torrenting that we did, which we did. And also, you know, they had that one search term for new custodians. And we ran that search term, by the way, against Mr. Lample, even though they didn't ask for him.

THE COURT: And -- but, I mean, if the corporation, at some point, was blocking torrenting and then unblocked torrenting, what does that -- I'm just trying to figure out, like, what does that -- does that change damages for the distribution? Like, what is it -- what does -- maybe it's a question for them --

MS. HARTNETT: I would be loath to argue their case for them. But I think that might be -- I mean, I just read what they say -- corporate imprimatur based on one nonlawyer's notes and an AIDC document.

THE COURT: Right.

MS. HARTNETT: I guess that's where they're trying to go with it.

THE COURT: Okay.

MS. HARTNETT: But, again, that was something we voluntarily produced, not because we were hiding it, but it was

not called for by the searches in the beginning. And then we have been erring on the side of doing the hyperlinks when they're responsive, per your order and Judge Hixson.

THE COURT: So -- and then the other -- I guess the other question I have that might be a dumb question is I seem to remember from the -- our last summary judgment proceeding -- right -- that the record -- it was indisputable that Meta, as a company, signed off on using shadow libraries at some point.

MS. HARTNETT: Well, I think that's an important distinction though, because there's a difference between whether the company made a decision to use the data sources and the method of the --

THE COURT: Torrenting.

MS. HARTNETT: Correct. And I think that was something your heard: Every time there's a download, that would be relevant to torrenting. That's not -- A, that's not right, because some downloading is not torrenting.

THE COURT: Right.

MS. HARTNETT: B -- right -- to the extent that there was an argument at summary judgment, it was about was there a corporate sign-off on the decision to use LibGen, not about torrenting.

THE COURT: And there was corporate sign-off on the decision to use LibGen, but we don't -- at least based on the record that I was looking at, at summary judgment, there was

nothing about -- or there was very little about the extent to which this stuff was downloaded by torrenting and, I guess, the extent to which the company was on board with that.

MS. HARTNETT: Yeah. And I'll just refer to the record on that in terms of how -- exactly what the evidence was before Your Honor. But, yes, that was the question before you then, was had Meta -- or to what extent did the evidence show Meta as an institution, as an entity, had signed off on the using of the datasets, not about torrenting.

THE COURT: Okay. All right.

Okay. Do you want to have the last word on any of this stuff? And I'll go back and think about it further.

MR. PRITT: Thank you so much, Your Honor.

Just to be clear, our motion for summary judgment previously was based on the torrenting of data from shadow libraries, because that's how Meta obtained almost all of its shadow library data, via torrent.

Certainly, the ML Hub issue of whether or not torrenting was approved is still relevant, to the extent Meta is going to raise a fair use defense to its peer-to-peer file sharing. I've asked Meta that, because its interrogatory response does not raise a fair use defense to that conduct.

Meta has said -- Meta's refused to disclaim raising fair use defense. And they said, instead, they might try to revise and serve a new interrogatory response. We don't know what

that might say. We have no idea if we need to take discovery into that.

We're not trying --

THE COURT: Is Meta raising a fair use defense to the distribution claim?

MS. HARTNETT: I'm sorry?

THE COURT: Is Meta raising a fair use defense to the distribution claim?

MS. HARTNETT: I think we haven't yet seen what their proof is, so we haven't made a determination about that yet.

THE COURT: Okay. That -- I think that you know whether you're planning on raising a fair use defense to their distribution claim or not.

MS. HARTNETT: I actually -- I honestly do not know that. I would have to consult with our client. But I think we actually have not decided how to oppose their evidence yet.

THE COURT: Okay.

MS. HARTNETT: We believe there's going to be -- we think it's going to be -- there will not be -- we expect -- and that's at least what happened with the expert last time, that there wasn't enough to show that any of the plaintiffs' works was torrented.

THE COURT: Okay.

MS. HARTNETT: So that is going to be the crux of it. I actually just have to consult with my client. I'm not

plugged into that work stream and don't want to make an inaccurate representation.

THE COURT: Okay.

MR. PRITT: I have asked that question since late August of 2024 several times and have not received a response.

This isn't just about the baseball arbitration discovery. If Your Honor recalls, the baseball arbitration discovery was about the 17,000 sequestered documents. In our proposal, we decided to make it specific to torrenting, because we had had no opportunity to conduct discovery. Because, if you recall, in January, when we proposed the limited reopening for uploading and distribution-related discovery, you denied our request and took Meta's, which was to not reopen discovery.

But, again, these are also relevant to some of the earliest RFPs served in this case. And when I mentioned our expert looking at the backups of the three main individuals -- their work laptops -- the reason I did that is because that's a short circuit of some of the discovery we already requested, where we asked for a search of the Lample files and servers. So I'm not expanding or creating an ocean of something. Instead, we're trying to focus on targeted searches.

In the RFAs -- I spoke with our expert just before this hearing. Again, respectfully, they're not expert discovery. They are literally the exact steps that we understand Ms. Wang to have conducted. And she's the only person who can answer

those. And so we provided those to Meta almost a month ago and are asking for it in advance of the deposition so we can shorten it and get actual answers that she's investigated and thought out to cut through things and cut the deposition in half.

That's all I have, Your Honor, unless you have any additional questions.

THE COURT: What about the issue of, you know, corporate signing off on torrenting? Well, if they -- if they were torrenting, and if, as part of that process, they shared your clients' works, why does it matter whether -- you know -- either one of two things happened; right? Either corporate was blocking it for a time and then stopped blocking it, or corporate was fine with it the whole time; right? And, either way, I don't think that would affect the company's liability for distribution; would it?

MR. PRITT: Well, it depends. There's some recent torrenting that Meta has said was just done by a contractor who worked at Meta that was -- they represent was not for AI training; it was for personal use. And they've seemed to be arguing some sort of affirmative defense to that torrenting that they're not responsible.

But if they are raising a fair use defense or some other affirmative defense to say, well, we never actually approved torrenting, all we approved was direct downloading, then it

would be highly relevant and our only way to respond to those affirmative defenses. So...

THE COURT: Well, I guess that begs the question that came up last time, which is, is bad faith relevant to fair use defense?

MR. PRITT: Yes. Whether bad faith is relevant to being able to invoke fair use at all or with respect to factor one. Yes.

And, again, I do want to say, we have asked for specific direct -- so if you go to 2022 and that downloading, where we only have 3.9 terabytes -- logs showing 3.9 terabytes of downloading -- but the documents indicate as much as 70 terabytes, we've identified specific file directories to Meta and said, can you look for these? And Meta will sometimes say, yes, we'll look for these two, but we're not going to look for the rest. We don't know why.

THE COURT: And why do you -- what do you want to discover? What do you want to learn from those file directories?

MR. PRITT: Whether or not the logs for the additional 65 terabytes of documents are there, because he left the company. And then, more recently, we learned that they --

THE COURT: But whether they're there or not, why would it matter?

MR. PRITT: Right. So if we're talking about a

statistical analysis of --

THE COURT: But it sounds like what you're saying is that we know that -- I can't remember the number of terabytes you're talking about -- but it sounds like you're saying, from the documents we already have, we know that he torrented X number of terabytes and we have a log for only Y number of terabytes. Why isn't that enough?

MR. PRITT: Circumstantial, but also because we're conducting summary judgment for class certification. So we're focused on the asserted works. So we need the logs to determine whether the asserted works were within those 65 terabytes of data that we don't have the logs for.

But we also learned, about a month ago, that Meta has a backup of Mr. Lample's work computer that he would have used to write code and to tell, based on a review of the computer, where -- what service he was using to actually download and torrent information. We have asked Meta, will you search that, and they've refused.

So there are certainly discrete categories of directories and work devices that we have pointed Meta to -- and I have a -- you know, this is why I said maybe it's best if our expert just goes in, because otherwise we're just doing this back-and-forth, where every time we ask for a -- have you looked at this directory, Meta says, we're not going to get into discovery on discovery. And we're still left wondering,

where are these missing files?

THE COURT: Were you wanting to --

MS. HARTNETT: I just jumped up, because we actually have looked for the additional Lample download, and we haven't been able to find it in the directory they're talking about, I believe. And the evidence that they're citing is allegedly that he downloaded more, but we only have the smaller part. It's a work chat where he was essentially saying, early, like a couple months before he left the company, that he couldn't find enough room if he were going to download more, not that he had downloaded more. So that's kind of the end of that trail. And no corroborating evidence that he actually did a larger download.

Just to be really clear, that's the level of on the limb that we are here. And we did look for the files, and we haven't found them.

MR. PRITT: Respectfully, Meta looked for two directories out of several directories. And they looked for two directories on a server called H2, which is owned by Penguin Computing. And they've refused to go to Penguin Computing to get the records that don't exist and we believe exist. And they've refused to look at any directories, other than two that we pointed out to them and asked them to look for.

MS. HARTNETT: That's because that's where that

individual did his torrenting.

THE COURT: All right. I'll give it some more thought and decide on the next step.

MR. PRITT: Thank you very much, Your Honor.

THE COURT: Thank you.

(Proceedings adjourned at 3:03 p.m.)

---oOo---

CERTIFICATE OF REPORTER

I certify that the foregoing is a correct transcript from the record of proceedings in the above-entitled matter.

DATE: Thursday, November 13, 2025

Kendra Steppler

Kendra A. Steppler, RPR, CRR

Official Reporter, U.S. District Court