July 31, 2025

E-Filed

The Honorable Thomas S. Hixson
United States District Court for the Northern District of California
San Francisco Courthouse, Courtroom E – 15th Floor
450 Golden Gate Avenue
San Francisco, CA 94102

Re: *Kadrey, et al v. Meta Platforms, Inc.*; Case No. 3:23-cv-03417-VC

Dear Magistrate Judge Hixson:

Plaintiffs to the above-captioned action ("Plaintiffs") and Defendant Meta Platforms, Inc. ("Meta") respectfully submit this joint letter brief regarding a dispute that has arisen as to whether Meta should be compelled to produce certain discovery, which is submitted to Your Honor pursuant to the stipulation and order regarding case scheduling proposals entered on July 28, 2025 (Dkt. 614).

I. PLAINTIFFS' POSITION

On February 7, 2025, months after the close of fact discovery, Meta informed the Court it sequestered and had not produced thousands of responsive documents (the "Sequestered Documents"). To remedy the resultant prejudice to Plaintiffs—which included belated disclosure of significant facts regarding Meta's torrenting activities—the Court directed the Parties to file competing proposals for additional discovery, one of which would be adopted in full through a "baseball-style arbitration." Dkt. 461. The Court adopted Plaintiffs' proposal. Dkt. 470. That proposal included, as relevant here: (1) an **RFP**[1] requiring Meta to produce documents showing how much and which data was uploaded during Meta's torrenting, including billing records for the torrent clients Meta used (Dkt. 467 at 2); and (2) a targeted **Search String**[2] similarly focused on evidence of Meta's torrenting, which Meta was to run across all existing custodians and "the work email and Workplace Chat of four *additional individuals* that the Sequestered Documents show were integral in torrenting from LibGen and Anna's Archive." *Id.* (emphasis added).

Meta has not complied fully with these court-ordered discovery obligations in two critical ways. *First*, Meta produced six months of AWS logs, but not logs for the rest of the court-ordered time period of three years. *Second*, Meta ran the court-ordered Search String only on the four newly-added custodians rather than all custodians. Plaintiffs request that the Court order Meta to produce this discovery.

These additional materials are of critical importance to the Parties' expert analysis and forthcoming summary judgment motions on Meta's torrenting. Judge Chhabria already noted that the record on Meta's uploading-related conduct requires further development (Dkt. 598 at 15, n.4) and remarked that Meta's uploading and related infringements are "important issues" that "need to be adjudicated on a full record to the extent possible." 7/11/25 Hr'g Tr. 16:25-17:1; *see also id.* 26:25-27:5 ("If you want to tee that [discovery dispute] up in front of the magistrate judge, you want to just get that resolved first, and then submit a stipulated schedule . . . that's fine.").

Plaintiffs initially flagged Meta's deficient productions in their March 24, 2025 opposition to Meta's administrative motion requesting leave for a rebuttal expert report and depositions (Dkt. 487 at 5). But the Court opted not to rule on the parties' discovery disputes at that time, noting that Plaintiffs could seek relief at a later stage if it became necessary. Dkt. 500 ("The Court declines at this late stage to address the plaintiffs' assertion (in their [487] opposition to Meta's request for leave to file a rebuttal report) that Meta is failing to comply with its discovery obligations. If Meta fails to produce material it is required to produce, the plaintiffs are free to seek relief under Rule 56(d) at the appropriate time (or to argue in their papers that they would be entitled to summary judgment if only Meta had produced documents that it needed to produce). Given the magnitude of this case, Meta is urged to take a 'when it doubt, produce it' approach.")." That time is now.

---

[1] "Documents sufficient to show the installation, usage, configuration, and operation of Torrent Clients to obtain Data from Online Databases, including the filename, size, file creation time, and file last modified date of the Data; configuration files or equivalent settings registries, bandwidth logs, work files, and support tickets; and Amazon Web Services invoices relating to Your data usage, storage, or transfer."

[2] (torrent* OR torent* OR leech*) AND (libgen OR "library genesis" OR lib-genesis OR library-genesis OR anna* OR AA OR "internet archive" OR IA OR z-lib! OR z lib! OR zlib OR z-library OR book! OR data OR cop* OR file*)

A. Missing AWS Logs

Plaintiffs' court-ordered **RFP** requested that Meta produce AWS invoices and data transfer logs. Dkt. 467-2 at 2. The RFP had a time period defined as follows: "The relevant date range for the Request for Production herein is January 1, 2022 to the present." *Id.* at 4.

On March 19, Meta made a limited production of AWS invoices and data transfer logs. For the first time in this case, Plaintiffs received direct,[3] quantitative evidence that Meta torrented massive amounts of copyrighted material via Amazon Web Services ("AWS") platforms in 2024. That data newly demonstrated the massive scale of Meta's file-sharing through Anna's Archive. But Meta only produced invoices and logs dating from April to October 2024, or the time period during which Meta claims it actively torrented via AWS in 2024. As Plaintiffs previously explained, the point of requesting the broader 2022-present time period was not just to verify Meta's own claims about its torrenting activities, but also to compare Meta's data-transfer usage during its self-proclaimed torrenting period against other periods. "If the invoices spiked around the periods that Meta was torrenting . . . it will provide additional evidence that uploading occurred" because "AWS bills its users based on, among other things, the amount of data uploaded out of AWS." Dkt. 467 at 3–4. Accordingly, Plaintiffs ask the Court to order production of the AWS logs for the full time period set forth in the RFP.

B. Incomplete Custodial Search of New Search String

Meta was obligated to run Plaintiffs' **Search String** over its 15 existing document custodians and "four *additional* individuals." Dkt. 467 at 2 (emphasis added). Instead, Meta ran the search *solely* over the four additional custodians' email and WorkChat, but not the original custodians' files. This yielded a production of about 150 documents, including several important torrenting-related documents that had previously evaded production.

Plaintiffs request that the Court order Meta to run the **Search String** across Meta's original 15 custodians and produce all responsive, non-privileged documents. Not only does the plain language of the discovery proposal make clear that it was to be run across existing *and* "additional" custodians, but it is nonsensical to read Plaintiffs' request any other way. Many of the significant torrenting-related documents to date originated from the existing custodians' files; yet, they were produced only by happenstance as they hit on other non-torrenting terms. Plaintiffs' **Search String** thus specifically aimed to target documents and communications that squarely discussed or captured torrenting-related evidence. Plaintiffs thus have good reason to believe this search will bear fruit and provide the full record that the Court requires to adjudicate the remaining issues.

[3] To date, the majority of the record on Meta's file-sharing of pirated copyrighted works came from WorkChats from Meta employees and assertions made by Meta's torrenting expert in her expert reports.

II. META'S POSITION

Fact discovery closed on December 13, 2024. In February 2025, Meta discovered that, due to an error by its vendor, about 1,350 responsive documents were inadvertently sequestered during fact discovery that had not been produced. (Dkt. 427, 427-1.) Meta immediately informed the Court of the issue and promptly produced these additional documents. (Dkt. 424, 427.) The Court observed that the document sequestration appeared to have been "an inadvertent mistake." (Dkt. 466, 25:23-26:2.) The Court directed both sides to simultaneously submit proposals for additional discovery to address the new documents, stating that it would "adopt the most reasonable proposal in a baseball style arbitration." (Dkt. 461; Dkt. 466, 40:4-41:19.) Plaintiffs' proposal argued for "limited additional discovery," represented as "significantly narrower" than a prior proposal rejected by the Court. (Dkt. 467, 1.) As relevant to this motion, Plaintiffs' proposal included a single new Request for Production No. 137 ("Request") and a single targeted search directed at four custodians who Plaintiffs believed were involved in the download of certain datasets using torrents. (*Id.*, 1-2.) But after the Court adopted Plaintiffs' proposal, and Meta provided the actual discovery Plaintiffs requested, Plaintiffs now seek to "reinterpret" their proposal to seek far more - despite originally representing to the Court that, if its proposed amendments to the Complaint were permitted (including allegations relating to torrenting), they "would not seek to initiate any new discovery" (Dkt. 376, 6-7). The parties have agreed, and the Court has ordered, that this dispute be referred to Judge Hixson. (Dkt. 614.) Meta has complied with its obligations with respect to this new discovery, and the Court should deny all requested relief.

A. Request for Additional AWS Records

Given that this Court previously ruled that Plaintiffs had not requested such documents during fact discovery (Dkt. 366 at 6-7), Plaintiffs added the Request, but now seek information about Amazon Web Services (AWS) logs far beyond what was requested about "Torrent Clients":

> Documents sufficient to show the installation, usage, configuration, and operation of Torrent Clients to obtain Data from Online Databases, including the filename, size, file creation time, and file last modified date of the Data; configuration files or equivalent settings registries, bandwidth logs, work files, and support tickets; and Amazon Web Services invoices relating to Your data usage, storage, or transfer.

(Dkt. 467-2 at 3 (underlining added).) Based on the plain language of the Request, "Amazon Web Services invoices relating to Your data usage, storage, or transfer" is simply an example of a type of data falling within the category of "[d]ocuments sufficient to show the installation, usage, configuration, and operation of Torrent Clients to obtain Data from Online Databases." Documents regarding the use of AWS for purposes ***other than*** this fall outside the scope of the Request.

AWS refers to "Amazon Web Services," a cloud computing platform offered by Amazon that companies can use to perform a variety of processing and computing tasks. (Dkt. 543-6, ¶ 104.) Meta has investigated the torrenting activity relating to Llama training datasets containing Plaintiffs' works, and based on that investigation, the ***only*** such torrenting activity utilizing AWS occurred in the April-October 2024 timeframe using several virtual machines (referred to as "instances"). Meta produced the AWS billing and usage data for these instances during this time period, thus producing all of the AWS records in its possession sought by the Request.

Plaintiffs did ***not*** request all AWS data from Meta across a multi-year period. As noted,

the ***only*** relevant torrent download for using AWS was in 2024, and Plaintiffs have no evidence to suggest otherwise. The 2024 AWS documents Meta produced thus target the activity from the ***actual AWS instances*** used for the torrent download, negating any need to compare AWS data over a longer time frame as Plaintiffs suggest. The other relevant torrent downloads did ***not*** use AWS but instead used Meta's own servers. For example, Meta produced thousands of log-related documents for a separate 2022 torrent download that was not performed utilizing AWS.

Finally, Plaintiffs' request for "the AWS logs for the full time period" is overly broad and impermissibly vague. Meta uses AWS pervasively across its operations to fulfill a broad range of compute needs, including content moderation and delivery across its social networking platforms (among many purposes having nothing to do with this case). Over the past several years, in fact, Meta has deployed millions of AWS instances, and at any given time, has tens of thousands in use. To the extent Plaintiffs seek data relating to *all* instances of AWS usage at Meta across the company from 2022 to the present, it would be unduly burdensome and disproportionate as it would require an enormous undertaking to collect, organize, and produce a massive volume of records for countless AWS instances unrelated to this case. The AWS data produced by Meta, from ***just*** the five AWS instances used for torrenting in 2024, was voluminous and time-consuming to collect; to extend that process across potentially millions of AWS instances would be overwhelmingly onerous, while offering nothing of evidentiary value because it would implicate instances and usage not tied in any way to Plaintiffs' claims. Plaintiffs' request should be denied.

B. Custodial Searches

In their March 5 proposal, Plaintiffs urged the Court to require Meta to perform "One Targeted Search" (Dkt. 467, at 2) and further identified the precise contours of the search, including the search term, the custodians, and the type of document sought to be searched:

> 4. **One Targeted Search.** This search[1] targets evidence of Meta's torrenting, including uploading, of copyrighted textual works from online databases of pirated works. Plaintiffs request this search include the work email and Workplace Chat of four additional individuals that the Sequestered Documents show were integral in torrenting from LibGen and Anna's Archive.[2]
>
> [1] (torrent* OR torent* OR leech*) AND (libgen OR "library genesis" OR lib-genesis OR library-genesis OR anna* OR AA OR "internet archive" OR IA OR z-lib! OR z lib! OR zlib OR z-library OR book! OR data OR cop* OR file*)
>
> [2] Those four individuals are Todor Mihaylov, Marie-Anne Lachaux, Xiaolan Wang, and Frank Zhang.

(*Id.* at 1-2.) Meta promptly complied and produced responsive documents from this search from these four custodians. Although not required to do so, Meta also produced from a fifth custodian, Guillaume Lample, a former employee who was also involved in relevant torrenting activity.

Plaintiffs now seek to rewrite their original proposal as follows: "Plaintiffs request this search include the work email and Workplace Chat of **Meta's fifteen custodians as well as** four additional individuals that the Sequestered Documents show were integral in torrenting from LibGen and Anna's Archive" (additions emphasized). But nowhere in their proposal did Plaintiffs request, or even suggest, that Meta apply this new search term to its original 15 custodians. Rather, Plaintiffs' use of the word "additional" merely identified to the Court that these were newly identified custodians that Plaintiffs claimed were "integral" in torrenting activities – not that Plaintiffs expected Meta to run the new search term across prior custodians.

Plaintiffs' motion also ignores that Meta already ran a torrent-related search on the original 15 custodians: (torrent AND (dataset OR "training data")). Plaintiffs also claim that "[m]any of the significant torrenting-related documents to date originated from the existing custodians' files," but for 14 of the original 15 custodians, there is no evidence they had any involvement in torrenting activity. The only original custodian involved in torrenting activity (Nikolay Bashlykov) was not involved in torrenting books, let alone Plaintiffs' works. (Dkt. 494, ¶¶ 4-6; Dkt. 543-6, ¶¶ 75-77, 85-87.) Plaintiffs' request for the Court to rewrite their search term proposal should be denied.

III. PLAINTIFFS' REPLY

With respect to the AWS Logs, what Meta produced is insufficient. To start, Meta doesn't argue the other Logs are irrelevant, only that—so Meta says—there aren't other "relevant torrent download[s]." Whatever that means, discovery is meant to prove or disprove a party's assertions. And even if the complete set of Logs support Meta's assertions, Plaintiffs would still need them to compare Meta's data transfers during the alleged non-torrenting periods. Indeed, Meta does not contest Plaintiffs' argument as to the relevance of the Logs for the non-torrenting periods. Meta's unsubstantiated burden claims also lack merit—Plaintiffs seek the same discrete AWS logs and invoices Meta already located and produced for part of the time period in the court-ordered RFP.

With respect to the custodial search, Meta ignores the plain meaning of the word "additional." More importantly, Meta misrepresents the original custodians' involvement in Meta's torrenting—for example, Meta claims the only original custodian implicated was Nikolay Bashlykov, but the first evidence of Meta's discussion of seeding came from the files of another custodian, Aurelien Rodriguez. *See* Dkt. 567-51. But even if Meta's claims were true, then simply running the Court-ordered search term over the complete set of custodians' files will result in few, if any, hits. Further, Meta's assertion that its preexisting search string (torrent AND (dataset OR "training data")) would have yielded all relevant documents is plainly wrong. That narrow search only captured documents mentioning "torrent" if they also contained the phrases "dataset" or "training data." This limits the universe to an unnatural subset —Meta employees do not torrent "training data" or "datasets," but rather "books," "files," or "data." Meta's search string thus produced just a few dozen documents mentioning "torrent*" before Meta ran Plaintiffs' court-ordered search over the four new custodians. The Court should order Meta to complete its search.

By: /s/ *Bobby A. Ghajar*

Bobby A. Ghajar
Colette A. Ghazarian
COOLEY LLP
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400
Facsimile: (310) 883-6500
Email: bghajar@cooley.com
cghazarian@cooley.com

Mark R. Weinstein
Elizabeth L. Stameshkin
COOLEY LLP
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000
Facsimile: (650) 849-7400
Email: mweinstein@cooley.com
lstameshkin@cooley.com

Kathleen R. Hartnett
Judd D. Lauter
COOLEY LLP
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111
Telephone: (415) 693-2071
Facsimile: (415) 693-2222
Email: khartnett@cooley.com
jlauter@cooley.com

Phillip Morton
COOLEY LLP
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004
Telephone: (202) 842-7800
Facsimile: (202) 842-7899
Email: pmorton@cooley.com

Angela L. Dunning
CLEARY GOTTLIEB STEEN & HAMILTON LLP
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4121

By: /s/ *Maxwell V. Pritt*

BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)
Joshua M. Stein (SBN 298856)
Margaux Poueymirou (SBN 356000)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jstein@bsfllp.com
mpoueymirou@bsfllp.com

Jesse Panuccio (*pro hac vice*)
Jay Schuffenhauer (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com
jschuffenhauer@bsfllp.com

Joshua I. Schiller (SBN 330653)
David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@bsfllp.com
jischiller@bsfllp.com

Interim Lead Counsel for Plaintiffs

Facsimile: (650) 849-7400
Email: adunning@cgsh.com

Attorneys for Defendant Meta Platforms, Inc.

ATTESTATION PURSUANT TO CIVIL LOCAL RULE 5-1(h)

I hereby attest that I obtained concurrence in the filing of this document from each of the other signatories. I declare under penalty of perjury that the foregoing is true and correct.

Dated: July 31, 2025

BOIES SCHILLER FLEXNER LLP

/s/ Maxwell V. Pritt
Maxwell V. Pritt

Attorneys for Plaintiffs