July 31, 2026

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 Judge Hixson:

The Parties respectfully submit this joint letter brief regarding Plaintiffs' motion to compel Meta to comply with the Court's orders to search and produce (1) command history for servers it used to torrent copyrighted works, and (2) responsive records from Meta's ML Hub. The parties have met and conferred several times, including on June 30, 2026, but were unable to reach resolution.

I. PLAINTIFFS' POSITION

a. Command History Files

In February 2025, Meta disclosed that it had inadvertently sequestered approximately 18,000 documents and had not reviewed or produced them. *See* Dkt. 420. Meta's subsequent productions revealed new facts about Meta's torrenting of copyrighted works. To decide how to address the prejudice caused by Meta's belated productions, Judge Chhabria ordered the parties to submit competing "baseball arbitration" proposals for further discovery. *See* Dkt. 461. The Court granted Plaintiffs' proposal, *see* Dkts. 467, 470, which included a single RFP seeking "Documents sufficient to show the installation, usage, configuration, and operation of Torrent Clients to obtain Data from Online Databases[.]" Dkt. 467-2. The RFP defined "Documents" to include "command-line history logs (including the .bash_history file in the user home directories of servers used for downloading data)[.]" *Id.*

More than a year later, Meta has not produced command history files from the servers used by three of Meta's four known torrenters to run the programs used to torrent copyrighted works from "pirated sites." Dkt. 654-7 at 10. Plaintiffs already know—from source code Meta has produced—that Meta engineers used computer programs to download torrent content in bulk. *See, e.g.*, Ex. A, Meta_Kadrey_00238502. But the source code tells Plaintiffs only how the torrenting code works, while the command history would indicate specifically which .torrent files—and thus which copyrighted works—Meta targeted for torrenting, and how much of each work Meta downloaded and uploaded.

Plaintiffs recently learned that Meta possessed but did not produce command history files as previously ordered (at Dkt. 470), because one of them hit on a search term that Meta was ordered to run, and Meta produced a command history file from a laptop used by one of the Meta engineers who torrented copyrighted works. Plaintiffs then met and conferred with Meta seeking production of all command history files as the Court had ordered, and Meta recently produced some command history files from the individual laptops of three Meta engineers who torrented copyrighted works. But these files only confirm that those three engineers did *not* use their laptops to run torrenting programs. Instead, they used various other servers, or "virtual machines," including Meta virtual computers and Amazon Web Services ("AWS") infrastructure. *See, e.g.*, Dkt. 727-1 at 2-3. Accordingly, the responsive command history files Meta was ordered to produce would come from *those* servers, not these three employees' laptops. Plaintiffs have identified many of these servers and identified the relevant file types for Meta to search, but Meta has not agreed to search for or produce the corresponding command history files.

Plaintiffs respectively request that the Court compel Meta to produce the command history files for each of the servers that it used to torrent copyrighted works, including Meta virtual computers, AWS servers, and any other virtual machines that Meta used for torrenting pursuant to Judge Chhabria's order at Dkt. 470. Plaintiffs further request that the Court order Meta to produce these records no later than July 31, 2026, so that Plaintiffs have sufficient time to analyze the data.

b. ML Hub Documents

In August 2025, Plaintiffs learned from a new production that Meta maintains a centralized database called the "ML Hub" that contains, among other things, an AI Data Catalogue ("AIDC").[1] Meta's counsel subsequently confirmed the ML Hub is a previously undisclosed non-custodial data source that had not been searched during the fact discovery period. *See* Dkt. 624 at 1 n.1.

[1] Meta produced a printout from its ML Hub entitled "AIDC_Anna's_Archive_Compliance." The AIDC is defined in another Meta document as "a unified experience for all data used in AI workflows," including evaluations and mitigations. Meta's counsel has since represented that the AIDC is no longer housed within the ML Hub, but Meta has not provided any information or produced any documents concerning where the AIDC is now located, when it was moved, or why.

Accordingly, in October 2025, Plaintiffs moved the Court for leave to take additional limited discovery, requesting that Meta search the ML Hub and its contents using an identified set of search terms. *See* Dkt. 635. Judge Chhabria granted Plaintiffs motion in full. Dkt. 647. Meta initially told Plaintiffs—and the Court—that it would produce the ordered documents by February 28, 2026. *See* Feb. 5, 2026 Hr'g Tr. 38:7-10. Meta informed Plaintiffs on March 10, however, that:

> Meta has not been able to construct a way to conduct searches of the non-AIDC portions of MLHub using Plaintiffs' search terms. Moreover, the type of information in the non-AIDC portions of MLHub does not appear to be related to Plaintiffs' remaining distribution claim. We are preparing materials for you to review about the non-AIDC portions of MLHub to help you understand why the non-AIDC portions would not be relevant to Plaintiffs' remaining distribution claim.

Ex. B.[2] Meta finally produced these documents—just 20 in total—last Friday. Notwithstanding Meta's representation that this discovery would "help Plaintiffs understand" why the ML Hub "would not be relevant to Plaintiffs' remaining distribution claim," the ML Hub appears squarely relevant to Plaintiffs' remaining claims because it is the central repository for Meta's dataset management. One document explains it "is a 'centralized machine learning hub' that was designed to streamline the entire ML Development lifecycle by serving as an integrated development environment for each development stage." Ex. C at -749. That same document indicates there is another section within the ML Hub entitled "Dataset Management in ML Hub." Ex. C at -751. Given that data acquisition is the first stage of model training, the scant records Meta has produced from the ML Hub suggest it likely contains records related to the acquisition and management of datasets, including those with copyrighted works that Meta torrented from pirate websites. Meta's recent limited production does not relieve Meta of its obligation to produce all documents in the ML Hub (since the time of the Court's order) that hit on the ordered search terms.

Regardless, Meta already made—and Judge Chhabria already rejected—Meta's arguments concerning the relevance of the ML Hub's contents to Plaintiffs' uploading claims. *See* Dkt. 637 at 4. Meta also did not argue then that it could not run search terms through "the non-AIDC portions of the ML Hub," but even if it still cannot construct the means to do so, Meta has had over *eight months* to search the individual files and pages in the ML Hub using the court-ordered search terms. Accordingly, Plaintiffs request that this Court order Meta to comply with the Court's November 2025 order and produce all responsive documents from the ML Hub by July 31, 2026.

II. META'S POSITION

Command History Files. Plaintiffs' request for "command history files" should be denied because it seeks information that falls outside the scope of RFP 137. As the Court may recall, RFP 137 was the subject of a previous letter brief and request by Plaintiffs that was rejected by the Court. More specifically, in August 2025, this Court rejected the Plaintiffs' efforts to expand RFP 137 to cover documents not requested by the language of the RFP, in that instance AWS records not related to torrenting activity. Dkt. 616. In making its ruling, the Court analyzed the language of RFP 137 and found that it sought three specific types of information about torrenting:

> (1) "information about the torrented data ('filename, size, file creation time, and file last modified date of the Data')", (2) "information about the torrent client ('configuration files or equivalent settings registries, bandwidth logs, work files, and support tickets')", and (3) "AWS invoices ('Amazon Web Services invoices relating to Your data usage, storage, or

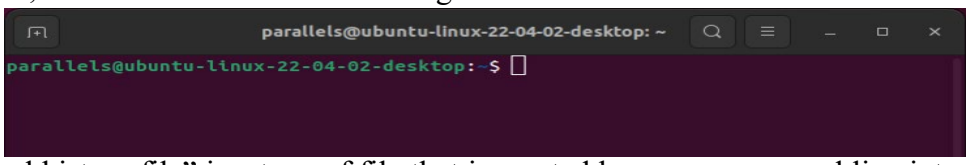
[2] To be clear, Plaintiffs do not have a single "remaining distribution claim." Plaintiffs assert a cause of action for direct infringement of their making available, reproduction, and distribution rights in connection with Meta's mass torrenting and uploading of pirated, copyrighted material. *See* Dkt. 702 ¶ ¶ 101-06. Plaintiffs also assert a claim for contributory infringement. *See id.* at ¶ ¶ 120-27.

transfer')."

Id. at 3. The Court further found that those three types of information were modified by the RFP's introductory language. *Id.* at 3 ("The only sensible way to read the RFP is that the introduction ('Documents sufficient to show the installation, usage, configuration, and operation of Torrent Clients to obtain Data from Online Databases, including . . .') modifies everything that follows. The semi-colons are used to separate three types of requested information about torrenting…").

Tellingly, Plaintiffs do not acknowledge the Court's prior order on RFP 137, nor attempt to tie its new request for "command history files" to any of the three specific types of information about torrenting that the Court found were requested by RFP 137. Plaintiffs instead pivot to a new argument, pointing to the RFP's broad definition of "Document" which states that it "is used in its broadest sense allowed by Federal Rule of Civil Procedure 34(a)," and then proceeds to list a number of exemplary types of Documents including "command-line history logs." *See supra* at 1; Dkt. 467-2, at 1. But simply because a command-line history log may qualify as a "Document" does not make it responsive to RFP 137 – it must still satisfy the remaining language of the RFP by falling into one of the three categories identified by the Court.

Plaintiffs speculate above that command-line history logs (or "command history file") might show which copyrighted works Meta targeted for torrenting, and how much of each work Meta downloaded and uploaded. But command history files would show no such thing. By way of background, one way to interact with UNIX-like computer operating systems is to enter textual commands (often typed by a user) into what is often referred to as a "terminal" or a "command line interface," such as that shown in the image below:



A "command history file" is a type of file that is created by some command line interfaces, such as BASH, that contains "the list of commands previously typed." *See* GNU BASH Reference Manual, Section 9.1 ("the shell provides access to the *command history*, the list of commands previously typed") (emphasis added), available at https://www.gnu.org/savannah-checkouts/gnu/bash/manual/bash.html#Bash-History-Facilities (last accessed July 29, 2026) (attached in pertinent part as Exhibit D). In other words, the command history files that Plaintiffs request may contain text that was entered into the command line interface to invoke a computer program or command, but they would not show the output or results of the computer program or command that gets executed in response. Critically, the "example" cited by Plaintiffs (Meta_Kadrey_00238502, Ex. A) in reply ***is not a "command history file"*** – it is a file Ms. Wang manually created listing scripts she used (which Plaintiffs characterize as "source code" above).

The documents sufficient to show what works were downloaded, and what information was downloaded and uploaded, has been provided to Plaintiffs through the voluminous productions Meta made many months ago including Meta's source code productions, gigabytes of file lists cataloging the downloaded files stored on Meta's servers, and hundreds of gigabytes of log files detailing Meta's downloading activities down to the minute. Below is a sample of one of those file lists, which shows which .torrent files were downloaded, the specific kind of information Plaintiffs assert they need the command history files for:

```
2024-05-28 22:21:37  132.3 KiB data/annas_archive/libgen/libgen_rs_non_fic/r_995000.torrent
2024-05-28 22:21:37  135.8 KiB data/annas_archive/libgen/libgen_rs_non_fic/r_996000.torrent
2024-05-28 22:21:37  133.0 KiB data/annas_archive/libgen/libgen_rs_non_fic/r_997000.torrent
2024-05-28 22:21:37  140.4 KiB data/annas_archive/libgen/libgen_rs_non_fic/r_998000.torrent
2024-05-28 22:21:37  135.6 KiB data/annas_archive/libgen/libgen_rs_non_fic/r_999000.torrent
```

Excerpt of Meta_Kadrey_00238482

The documents that are already in Kadrey's possession are more than sufficient to show which torrent files are at issue, and provide the information available to Meta about how much data Meta

downloaded and uploaded.[3] The Court should deny Plaintiffs' request for command history files.

MLHub

Plaintiffs' motion demanding that Meta conduct additional searches of MLHub (beyond those already conducted in AIDC) will serve no purpose other than to engage a needless search of a resource that does not contain information relevant to Plaintiffs' remaining distribution claims.

MLHub is an internal Meta website that contains links to a collection of tools that are used for machine learning development. Lindsay Decl. ¶ 4 (Ex. E). Machine learning extends much broader than the Llama generative AI models at issue in this case, as it includes other engineering activities like content suggestions or ads ranking on Meta's Instagram and Facebook apps. *Id.* at ¶¶ 4-5. MLHub is principally used by engineers working on these types of "classic" machine learning activities and that do not work on the Llama generative AI models. *Id.* at ¶ 5.

The only portion of MLHub that is potentially relevant to LLM development and Plaintiffs' remaining distribution claims is the dataset management information contained in the previously searched AIDC (the "AI Data Catalog"). *Id.* at ¶¶ 5-6. AIDC is the only resource within (or now linked from) MLHub that provides information about datasets used for training. *Id.* at ¶ 6. AIDC has been used in the past for cataloging information about datasets for potential use in LLM training, including tracking approvals by privacy, policy and legal teams. *Id.* at ¶ 6. As acknowledged by Plaintiffs, Meta has already searched for and produced responsive documents from AIDC based on the search terms allowed by the Court's Dkt. 647 order.

Plaintiffs' only argument about the purported relevance of MLHub is a citation to an internal webpage produced by Meta that references a webpage titled "Dataset Management in MLHub." But that webpage explains that this dataset management functionality in MLHub is AIDC, which Meta already undertook extensive efforts to search. *Id.* at ¶ 7; Ex. 1. That search required significant engineering work to construct a system to conduct keyword searches of AIDC because it lacked the native capability to conduct the searches that Plaintiffs were permitted to pursue in Dkt. 647. Those efforts were then followed with significant manual effort to collect the AIDC materials for review and production.

Even if MLHub contained information outside of AIDC that was relevant to Plaintiffs' remaining distribution claims, Meta does not have a readily available, reliable mechanism to search across all of the tools in MLHub using keyword searches. *See id.* at ¶ 8. MLHub is a collection of links to tools and is not itself a database or data repository. *Id.* Accordingly, no keyword search functionality exists within MLHub, nor would it make sense for Meta to build such functionality. *Id.* Each of the tools accessible via MLHub are subject to their own tool-specific access controls. *Id.* Access to MLHub does not allow access to all tools or information available through those tools, unless explicitly granted. *Id.* It would not be proportional to the needs of the case to require Meta to search the non-AIDC portions of MLHub as it would necessitate Meta undertaking the substantial burden of engineering means to search many individual tools that are not related to LLM development and do not have information related to Plaintiffs' remaining distribution claims.

Finally, Plaintiffs are incorrect in their assertion that the Court "rejected" Meta's relevance or burden arguments. In allowing the discovery that included the MLHub and AIDC searches, the

[3] Additionally, Plaintiffs' interpretation of this RFP as covering information like "command history files" is further belied by its motion for additional discovery that it filed in October 2025. Dkt. 636-1. In that Motion, Plaintiffs asked the Court to order Meta to conduct searches of personal computers and servers for evidence about torrenting related to one individual, Mr. Lample, to obtain materials like command history files. Dkt. 636-1 at 5. Tellingly, Plaintiffs did not argue that those searches were covered by the previously served RFP 137; it treated them as entirely new discovery that the Court had to grant Plaintiffs permission to pursue. In an effort to provide transparency to Plaintiffs and the Court regarding these activities, Meta later voluntarily expanded those searches to three other Meta personnel.

Court was clear in its order that "[t]he Plaintiffs may pursue the discovery" described in their motion, not that Meta had no ability to raise any objections to relevance or burden. Dkt. 647.

Conclusion

Plaintiff's requests for "command history files" and searches of MLHub should be denied in their entirety.

III. PLAINTIFFS' REPLY

1. "**Command history files**" fall squarely within the discovery sought by RFP 137.[4] Plaintiffs defined "Documents" to include "command history files" because these files are quintessential torrenting evidence. They reveal how torrent downloads were configured and stored and thus are "documents sufficient to show the . . . configuration[] and operation" of Meta's Torrent Clients. Meta's productions of other torrenting records do not relieve it from producing this unique data. Meta could have tracked its torrenting activity but chose not to—so Plaintiffs instead must piece together Meta's torrenting and distribution through different evidentiary sources. The directory listings Meta references above, for example, do not contain information about the specific software configurations Meta used to torrent; the command history files do. And while the directory listings contain information about where Meta stored torrented data, the listings are incomplete because Meta deleted vast amounts of data it torrented. While Meta relies on a generic example of command history to downplay its relevance, examples from Meta's productions show the opposite:

```
fair_data/projects/annas_archive/download_spark.py  \
--dataset_name="zlib" \
--input_path="s3://fairspark-data/data/annas_archive/zlib/zlib.parquet" \
--output_dir="s3://fairspark-data/data/annas_archive/zlib/data/" \
--status_output_dir="s3://fairspark-data/data/annas_archive/zlib/" \
```

Excerpt of Ex. A. This language, which appears to be commands intended for the command line that tell us the specific torrent files used, where the torrented content was placed, and where the completion status of this torrenting resides, should be analogous to the content in the command history files Plaintiffs' seek. Finally, contrary to Meta's claims, the burden is minimal—it comprises a handful of discrete files on specific servers that Plaintiffs' expert estimates Meta can identify in mere minutes. Meta should thus be compelled to identify and produce this highly relevant and previously ordered discovery. *See also* Dkt. 500 (Chhabria, J.: "Meta is urged to take a 'when it doubt, produce it' approach.").

2. Meta's relevance and burden arguments about searching the **ML Hub** are misplaced. This is a motion to compel compliance with an existing order, not a new dispute. Judge Chhabria ordered Meta to search the ML Hub using search terms, Dkts. 635 (Mot.) & 647 (Order), over Meta's relevance and burden objections, Dkt. 637 at 4 (Opp.), and after discussion at a hearing on the motion that the ML Hub discovery goes beyond distribution. *See generally* 11/10/2025 Hr'g Tr. Judge Chhabria also already gave Meta the option of alleviating any burden in conducting the ordered searches by having Plaintiffs' expert inspect the relevant sources "in lieu of [Meta] running the requested searches itself[.]" Dkt. 647 at 1. *See also* 5/22/2026 Hr'g Tr. 9:3-5 (Chhabria, J.: "when I hear concerns about burden [from Meta] in a case like this, I get concerned"). Further, while the Lindsay Declaration (Ex. E) states the ML Hub has information beyond AI models, that is irrelevant to the fact that it does include relevant information. *See* Dkt. 637 at 4 (Meta: "ML Hub is an internal Meta tool with information about AI models and experiments."). The Declaration also just states that the AIDC portion of the ML Hub is used for "cataloging" datasets, not that it's the only part of the ML Hub relevant to the millions of copyrighted works Meta torrented and compiled into datasets, and what Meta did with them. The Court should thus order Meta to comply with Judge Chhabria's order that it search the ML Hub or let Plaintiffs inspect it.

[4] Meta's characterization of this Court's 2025 Discovery Order on RFP 137 is misleading and irrelevant: that dispute concerned the responsiveness of AWS logs for non-torrenting periods, and this Court held the RFP was limited to torrenting-focused requests. *See* Dkts. 615, 616.

By: /s/ *Phillip Morton*

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*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com

David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@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:

BOIES SCHILLER FLEXNER LLP

/s/ Maxwell V. Pritt
Maxwell V. Pritt

Attorney for Plaintiffs