

December 20, 2024

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. 23-cv-03417-VC (TSH)

Dear Judge Hixson:

Plaintiffs and Defendant Meta Platforms, Inc. (“Meta”) jointly submit this letter brief regarding Plaintiffs’ request for an order compelling Meta’s production of the documents and data described herein. The parties met and conferred on December 17, but were unable to reach a resolution.

I. PLAINTIFFS' STATEMENT

In his recent deposition, Meta research engineer Nikolay Bashlykov testified to both the existence and the location of key data underlying the Llama Models' training datasets of copyrighted works and copyright mitigation functions. Subsequent communications with Meta revealed that *none* of this data was produced to Plaintiffs or their experts. It should be. Plaintiffs thus respectfully request that the Court order Meta to produce the following three categories of data.

A. [REDACTED] Data

Meta's [REDACTED] has become a topic of much attention in this case. *See, e.g.*, Dkt. Nos. 300; 308; 321; 335; 342. Once Meta's witnesses began testifying about this practice, it became clear that Meta's copyright infringement was far more brazen than Plaintiffs previously knew or assumed. Notably, two Meta 30(b)(6) witnesses testified that Meta [REDACTED]

[REDACTED]. The full extent of Meta's [REDACTED] formatively bears on Meta's intentional copying and use of pirated books and awareness that this conduct was legally problematic given Meta's efforts to prevent the public from being able to trace its [REDACTED] back to Meta IP addresses and Facebook servers. Meta_Kadrey_00204223 (Bashlykov: "[REDACTED] [REDACTED]

Meta's [REDACTED]-related data is thus directly relevant to Plaintiffs' copyright infringement claim because it reflects some of the copyrighted data that Meta downloaded from the shadow/pirated libraries at issue in this case, and it is also evidence of Meta *distributing* this copyrighted data without consent from the actual copyright holders, which is an independent infringing act. To fully assess the scope of Meta's [REDACTED], Plaintiffs asked Meta to produce its [REDACTED]—data that reflects how much Meta downloaded and from where, and how much Meta [REDACTED] (i.e., reuploaded) to the internet. This data is unquestionably responsive to Plaintiffs' RFP 85 ("All Documents and Communications related to the decision to use [REDACTED] Systems to acquire data for training Llama Models") and RFP 119 ("All Documents and Communications, including source code, relating to the processing of copyrighted material used in training Llama Models, including storage and deletion of copyrighted material."). While Meta has produced *other* responsive [REDACTED]-related documents, Meta refuses to produce the actual data about what was [REDACTED]. That is fundamental discovery and the Court should order its production.

B. Supervised Fine-Tuning Data

Mr. Bashlykov also testified that Meta stores its "supervised fine-tuning data" for its Llama models on [REDACTED]. Ex. B at 144-146. Plaintiffs have observed what appeared to be gaps in Meta's mitigation data productions, and Plaintiffs now know from Mr. Bashlykov that this set of fine-tuning data exists within a discrete data location that has not been produced. Plaintiffs' RFPs 118 and 119 (already subjects of a pending discovery letter, Dkt. No. 308) also cover this specific fine-tuning dataset. For the same reasons Plaintiffs explained in that letter, Meta's supervised fine-tuning data is relevant. That data regulates Llama by (1) training the model to identify when copyrighted material has been emitted and (2) preparing alternative answers when copyrighted emissions occur. The supervised fine-tuning data consists

of copyrighted works themselves: in short, the model is fine-tuned to say, “Don’t emit this.” Thus, not only does Meta’s supervised fine-tuning data itself consist of Plaintiffs’ and putative class members’ works, but whether Llama models frequently regurgitate copyrighted material unless fine-tuned also bears on Meta’s fair use argument that Llama models’ outputs are “transformative.”

There is no debate that the supervised fine-tuning data exists. Mr. Bashlykov testified to where it’s located. And it is clearly responsive. The Court should therefore compel its production.

C. Llama 4 and 5 Training Datasets

This Court already held that Llamas 4 and 5 are not analytically different from Llamas 1-3. *See* Dkt. No. 279 at 4 (“Llama 4 is relevant to this case, notwithstanding that it is still under development.”); Dkt. No. 315 at 7 (“Llama 5 is relevant”). In its order addressing the parties’ dispute over the definition of “Shadow Datasets,” the Court also stated that “since Meta has disclosed all the datasets that were used from Llama 1, 2, and 3, Plaintiffs can tell Meta which of those are Shadow Datasets.” Dkt. No. 315 at 8. Meta’s representation that it has fully produced the training datasets for Llamas 1-3 resolves *that* issue. But Meta has not done the same for Llamas 4 or 5, and indeed refuses to do so. Instead, Meta has only produced “documents sufficient to show various other datasets that have been or are being considered for use with future models.” *Id.* at 7.

The problem is that Meta’s witnesses have testified that [REDACTED]. For example, the evidence makes clear that over the last several months, Meta has [REDACTED], which had numerous domains seized by the FBI in recent years.¹ Moreover, Meta has begun to source new copyrighted works for its [REDACTED]. Yet, Meta has not produced any of this data, and Plaintiffs still do not possess the full Llama 4 (or Llama 5) training datasets despite their relevance.

This missing training data plainly exists. When Plaintiffs emailed Meta to confirm whether it had produced the entire training datasets for Llamas 4 and 5, Meta responded with a bizarre objection that the request does not correspond to any RFPs. Meta’s response clearly indicates that Meta possesses more Llama 4-5 training data; if Meta did not, it would have said that all responsive data was produced. Further, Meta’s contention is untrue: the Llama 4 training dataset is responsive to multiple RFPs. RFP 81 asks for “All Documents and Communications related to the decision to use Shadow Datasets for training Llama Models,” which would encompass the actual training data pulled from the Shadow Datasets (certainly the actual datasets are “related to” the decision to use those datasets). Dkt. 294-1 at 8-9. RFPs 6-12 also cover all documents and communications with enumerated Shadow Datasets concerning training data. And finally, the fact that Meta purportedly produced documents “sufficient to show” the datasets under consideration for use with future models, Dkt. No. 315 at 7, is a concession of relevance—Meta would not have produced “sufficient to show” documents that fall outside its interpretation of Plaintiffs’ discovery requests.

Accordingly, Plaintiffs request an order compelling production of Meta’s complete training datasets used for Llama 4 (and Llama 5 if any datasets exists), and at minimum, the training data subsets derived from the sources listed in Plaintiffs’ definition of “Shadow Datasets”: Books3, Z-

Library (aka B-ok), Library Genesis, Bibliotik, Anna s Archive, and The Pile. Dkt. No. 294-1 at 4.

META'S STATEMENT

Documents Related to Alleged “██████████” In this motion, Plaintiffs distort deposition testimony to once again demand documents they did not ask for in discovery and that are not relevant to the lone remaining copyright infringement claim—namely, Meta’s alleged “██████████

██████████. Plaintiffs did not ask for this information in their document requests. And Plaintiffs readily admit that Meta already has produced documents regarding discussions of alleged ██████████ within Meta, including documents regarding any alleged decisions to ██████████. There is nothing to compel here.

Plaintiffs cite two RFPs in their brief, Nos. 85 and 119, but neither calls for any of the documents Plaintiffs now demand. RFP 85 is specifically directed to decision-making about the alleged use of ██████████ “All Documents and Communications related to the *decision to use* ██████████” Ex. D (RFP 85). Contrary to Plaintiffs’ assertion, RFP 85 does not mention and cannot plausibly be read to request “██████████” much less any alleged ██████████

The Court has already ruled that RFP 119 does not cover requests for documents about ██████████. ECF 351 at 2-3 (“The Court does not see how ██████████ is responsive to this RFP, which is about the processing of data, not its acquisition”). Consistent with the Court’s prior ruling, as much as Plaintiffs would like to use this RFP as a catchall for anything they might want in the moment, it makes no reference to ██████████ or other means of *acquiring* data for training the Llama models. Rather, it seeks documents relating to the *processing* of copyrighted material used in training, including the storage and deletion of such material.

If Plaintiffs had intended to request production of ██████████, there was a simple way to do that. Plaintiffs made no such request. Indeed, the ██████████ sought by Plaintiffs are the types of materials that the parties agreed in the ESI order are not proportional to the needs of the case. ECF 101 at 7. ██████████ would at least qualify as “[s]erver, system, or network logs” and “[o]n-line data such as temporary internet files, history, cache, cookies, and the like,” which the ESI order exempts from production in this case. ECF 101 at 7 (¶ 8.C.iii & vi).

Supervised Fine-Tuning Data. Plaintiffs admit that they have already asked for this material in their pending motion on RFPs 118-119, which was heard at the December 19, 2024 hearing, with the Court asking for supplemental briefing with a narrower “ask” from Plaintiffs. Even if the Court were willing to entertain yet another motion on these same RFPs (it should not), neither supports Plaintiffs’ request for production of “fine-tuning data.”

RFP 118 seeks documents “relating to any efforts, attempts, or measures implemented by Meta to prevent Llama Models from emitting or outputting copyrighted material.” This RFP is directed to documents describing or showing “efforts, attempts, and measures,” which Meta has produced. It does *not* ask for training data of any kind.

Likewise, the Court has already ruled that RFP 119 does not cover datasets or copies of copyrighted works. ECF 351 at 2-3 (noting that RFP 119 “did not ask for” “datasets that include

Plaintiffs' copyrighted works"). As the Court knows, RFP 119 seeks "All Documents and Communications, including source code, relating to the *processing* of copyrighted material used in training Llama Models, including storage and deletion of copyrighted material." As the Court has previously recognized, Plaintiffs cannot morph a request for documents describing the processing, storage, and deletion of copyrighted material into a request for post-training data they never asked for. Additionally, as with the prior motion, Plaintiffs again baldly assert that fine-tuning data "bears on" the issue of transformativeness for fair use, but again provides no explanation why or how it is connected to the issues in dispute.

Finally, Plaintiffs' citations to the Bashlykov deposition do not support their argument that any of this material is relevant here. The muddled questioning does not evidence any specific tie between the fine-tuning data Mr. Bashlykov is talking about and any copyrighted works, let alone the Plaintiffs' works, nor any connection to any alleged copyright mitigation efforts. *See* Ex. B at 144-46. Plaintiffs' motion on supervised fine-tuning data should be denied.

Llama 4 and 5 Training Datasets. Plaintiffs did not ask for any training datasets in discovery other than the datasets for Llamas 1-3. Indeed, Plaintiffs knew how to specifically ask for training datasets; their first three RFPs served on December 27, 2023 specifically asked for "The Training Data for Llama [1/2/3]."² Ex. E (excerpt of RFPs 1-3). And there is no dispute that Meta produced that data in the case, subject to compromises between the Parties, due to the burden of collecting these datasets, to limit the data for Llama 3 to datasets relating to books, similar to Plaintiffs' requested "minimum relief" here. Plaintiffs' RFPs specifically asked for training data for Llama 1, 2, and 3 *only*. They never asked for the training data for Llama 4 or Llama 5.

In the absence of specific RFPs directed to Llama 4 or 5 training data, Plaintiffs cite RFPs 6-12 and 81. To start, RFPs 6-12 (like RFPs 1-3, which Plaintiffs notably do not reference) are Existing Written Discovery served nearly a year ago, and any motion practice directed to those RFPs was required to be filed no later than Nov. 8, 2024, as extended by Judge Chhabria following Plaintiffs' failure to comply with the Court's earlier deadline. ECF 253. Plaintiffs' reliance on these RFPs as a basis for its current motion to compel disregards Judge Chhabria's orders. Moreover, these Requests seek communications with various alleged "organizations," not the datasets themselves.

Plaintiffs also cite to RFP 81, but that RFP does not support the production of "all" Llama 4 and Llama 5 training data. Like RFP 85 discussed above, RFP 81 is specifically directed to materials about *decision-making* surrounding the alleged use of "Shadow Datasets," not the datasets themselves: "All Documents and Communications *related to the decision to use* Shadow Datasets for training Llama Models" And even if this RFP could be distorted to cover actual training data (which it should not), Plaintiffs do not even attempt to explain how RFP 81 might require production of "Meta's complete training datasets used for Llama 4 (and Llama 5 if any datasets exists)", which Plaintiffs concede are not limited to so-called "Shadow Datasets."

Finally, producing all training data (let alone multiple copies) for Llama 4 would be exceptionally burdensome and not proportional to the needs of the case. Meta has investigated the burden involved and just the datasets in Plaintiffs' requested "minimum" relief of identified "Shadow Datasets," when exported, would result in [REDACTED] of data across [REDACTED] data tables, comprising [REDACTED] of data [REDACTED]. Meta estimates that, assuming

² Unlike other requests asking about "Llama Models" generally, where the Court found Llama 4 and 5 to be relevant, these Requests seek documents re these specific Llama models only.

there are no unanticipated issues, for just [REDACTED] of the data tables it would take at least five weeks to complete the export process alone. The export process for those [REDACTED] tables is anticipated to take at least five weeks because [REDACTED]

[REDACTED] It will also take several weeks of machine processing time to export the data into a format that can be produced for review. Export of the other [REDACTED] data tables in the “minimum” relief requested by Plaintiffs is estimated by Meta to take at least ten weeks. Had Plaintiffs’ timely and explicitly requested this, as they did for Llama 1-3 (RFPs 1-3), Meta could have produced this data, but this eleventh-hour ask after the close of discovery makes this infeasible. Moreover, all training datasets that may be used for Llama 4 (and to the extent known for Llama 5) would constitute [REDACTED] [REDACTED] data tables and it is unclear that this amount of data could even be exported, and if so, how long it might take.

Plaintiffs’ demands for just the alleged “Shadow Datasets” would be incredibly burdensome exercise that would provide minimal if any relevant information for the issues in this case, as Plaintiffs’ works, to the extent they are included in the training data, are a miniscule portion of the overall training data set, and Meta has already admitted that text from each of the Plaintiffs’ books was included in the Books3 dataset used to train Meta’s Llama models. (See amended responses to RFAs 3–6, ECF 352, Ex. A.) Meta has also already produced documentation to Plaintiffs that identifies the datasets that are being used for ongoing Llama 4 training. Llama 5 remains at early stages of planning and it is not yet known what datasets will be used for training at this time.

III. PLAINTIFFS’ REPLY

With respect to another dispute, this Court did not order production of [REDACTED]. Dkt. 351 at 2–3. Yet, as is more patently relevant here, RFP 119 concerns “processing.” Indeed, [REDACTED] is not acquisition but *processing*. A party must *process* [REDACTED]. Thus, [REDACTED] is also an act of data processing.³ Understanding how Meta processed this data will speak to what was copied and how. It will also demonstrate how, to whom, and what exactly was distributed. Similarly, [REDACTED].

With respect to Meta’s decision to [REDACTED] (RFP 85), *how* Meta [REDACTED] [REDACTED] because some methods would have more readily facilitated or concealed Meta’s infringement. The more we know about that, the more we can understand why Meta *decided* to use which [REDACTED] (relevant, at minimum, to willfulness and bad faith). Here, Meta has produced fewer than 30 documents discussing [REDACTED]. Yet, Meta deponents have testified that [REDACTED].

Meta should also produce the Shadow Library training datasets used for training Llamas 4 and 5—much of which it copied well before the launch of Llama 3. “Communications” with these Libraries (including direct downloads which are, by definition, communications) encompasses the training data Meta copied from them. Meta cannot claim burden simply because it stole too much data.

³ Ask Meta.AI, “Does [REDACTED] involve data processing,” and it replies, “Yes, [REDACTED] involves data processing in various ways.” See *ME2 Prods., Inc. v. Bayu*, 2017 WL 5165487, at *1-2 (D. Nev. Nov. 7, 2017) (re: processing); *New Sensations, Inc. v. Does 1-426*, 2012 WL 4675281, at *2 (N.D. Cal. Oct. 1, 2012) (Corley, J.) (same); see also [REDACTED]

By: /s/ Bobby 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)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jstein@bsfllp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@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: December 20, 2024

BOIES SCHILLER FLEXNER LLP

/s/ Maxwell V. Pritt _____

Maxwell V. Pritt

Reed Forbush

Jay Schuffenhauer

Attorneys for Plaintiffs

EXHIBIT A

1 COOLEY LLP
BOBBY GHAJAR (198719)
2 (bghajar@cooley.com)
COLETTE GHAZARIAN (322235)
3 (cghazarian@cooley.com)
1333 2nd Street, Suite 400
4 Santa Monica, California 90401
Telephone: (310) 883-6400
5 Facsimile: (310) 883-6500

6 MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
7 KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
8 JUDD LAUTER (290945)
(jlauter@cooley.com)
9 3175 Hanover Street
Palo Alto, CA 94304-1130
10 Telephone: (650) 843-5000
Facsimile: (650) 849-7400

11 *Counsel for Defendant Meta Platforms, Inc.*

12
13 **UNITED STATES DISTRICT COURT**
14 **NORTHERN DISTRICT OF CALIFORNIA**

15 RICHARD KADREY, et al.,
16 Individual and Representative Plaintiffs,
17 v.
18 META PLATFORMS, INC., a Delaware
19 corporation;
20 Defendant.

Case No. 3:23-cv-03417-VC

**DEFENDANT META PLATFORMS, INC.’S
OBJECTIONS AND RESPONSES TO
PLAINTIFFS’ FIRST SET OF REQUESTS FOR
PRODUCTION**

Trial Date: None
Date Action Filed: July 7, 2023

1 Meta objects to this Request to the extent that it implies that Meta is responsible for having
2 “assembl[ed]” the Gutenberg dataset.

3 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
4 needs of the case to the extent that it seeks “all” documents and communications concerning the
5 subject matter of the Request, including documents with limited, if any, relevance to Plaintiffs’
6 copyright infringement allegations and Meta’s defenses thereto.

7 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
8 Protective Order and ESI Order, Meta will conduct a reasonable search for and produce, or, if
9 technically infeasible or impractical to produce, identify public sources of or permit inspection of
10 a copy of the “Gutenberg” dataset referenced in the paper “LLaMA: Open and Efficient Foundation
11 Language Models.” Meta will also conduct a reasonable search for and produce non-privileged,
12 relevant documents, if any, in its possession, custody, or control concerning the selection of said
13 dataset for use in training any Meta Language Models (as construed above). Following the
14 identification of appropriate search terms and custodians by Plaintiffs pursuant to the ESI Order to
15 be used to search for email and electronic communications, and subject to the foregoing objections
16 as well as Meta’s objections to said custodians and search terms, Meta will also search for and
17 produce non-privileged, relevant communications responsive to this Request, if any, in accordance
18 with the limitations set forth in the ESI Order. Any such production shall be on a rolling basis, and
19 the timing of any such production shall be governed by terms of the Scheduling Order (ECF No.
20 87).

21 **REQUEST FOR PRODUCTION NO. 6:**

22 Documents and Communications to, from, or with Project Gutenberg Concerning Training
23 Data.

24 **RESPONSE TO REQUEST NO. 6:**

25 Meta incorporates by reference its objections and definitions above, including to the terms
26 “Training Data” and “Communications.” Any electronic communications, including email and
27 documents attached thereto, must be requested pursuant to the ESI Order.
28

1 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
2 Communications to, from, or with” to the extent it is seeking materials other than those governed
3 by the procedure for requesting email and electronic communications under the ESI Order, and
4 Meta interprets the Request as seeking such email and electronic communications and Documents
5 attached thereto. This Request is also vague and ambiguous as to the term “Project Gutenberg,”
6 which is undefined. In the context of this Request, Meta will construe this phrase to mean the
7 organization referenced here: <https://www.gutenberg.org/>.

8 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
9 needs of the case to the extent that it is interpreted to seek all documents and communications
10 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
11 and disproportionate to the needs of the case to the extent it seeks documents and communications
12 in the possession, custody, or control of, and/or concerns a third-party.

13 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
14 Protective Order and ESI Order, following the identification of appropriate search terms and
15 custodians by Plaintiffs pursuant to the ESI Order to be used to search for email and electronic
16 communications, and subject to the foregoing objections as well as Meta’s objections to said
17 custodians and search terms, Meta will search for and produce non-privileged, relevant
18 communications responsive to this Request, if any, in accordance with the limitations set forth in
19 the ESI Order. Any such production shall be on a rolling basis, and the timing of any such
20 production shall be governed by terms of the Scheduling Order (ECF No. 87).

21 **REQUEST FOR PRODUCTION NO. 7:**

22 Documents and Communications to, from, or with Library Genesis (aka LibGen)
23 Concerning Training Data.

24 **RESPONSE TO REQUEST NO. 7:**

25 Meta incorporates by reference its objections and definitions above, including to the terms
26 “Training Data” and “Communications.” Any electronic communications, including email and
27 documents attached thereto, must be requested pursuant to the ESI Order.
28

1 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
2 Communications to, from, or with” to the extent it is seeking materials other than those governed
3 by the procedure for requesting email and electronic communications under the ESI Order, and
4 Meta interprets the Request as seeking such email and electronic communications and Documents
5 attached thereto. This Request is also vague and ambiguous as to the term “Library Genesis,”
6 which is undefined. In the context of this Request, Meta will construe this phrase to mean the
7 organization referenced here: <http://libgen.li/>.

8 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
9 needs of the case to the extent that it is interpreted to seek all documents and communications
10 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
11 and disproportionate to the needs of the case to the extent it seeks documents and communications
12 in the possession, custody, or control of, and/or concerns a third-party.

13 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
14 Protective Order and ESI Order, following the identification of appropriate search terms and
15 custodians by Plaintiffs pursuant to the ESI Order to be used to search for email and electronic
16 communications, and subject to the foregoing objections as well as Meta’s objections to said
17 custodians and search terms, Meta will search for and produce non-privileged, relevant
18 communications responsive to this Request, if any, in accordance with the limitations set forth in
19 the ESI Order. Any such production shall be on a rolling basis, and the timing of any such
20 production shall be governed by terms of the Scheduling Order (ECF No. 87).

21 **REQUEST FOR PRODUCTION NO. 8:**

22 Documents and Communications to, from, or with Z-Library (aka B-ok) Concerning
23 Training Data.

24 **RESPONSE TO REQUEST NO. 8:**

25 Meta incorporates by reference its objections and definitions above, including to the terms
26 “Training Data” and “Communications.” Any electronic communications, including email and
27 documents attached thereto, must be requested pursuant to the ESI Order.
28

1 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
2 Communications to, from, or with” to the extent it is seeking materials other than those governed
3 by the procedure for requesting email and electronic communications under the ESI Order, and
4 Meta interprets the Request as seeking such email and electronic communications and Documents
5 attached thereto. This Request is also vague and ambiguous as to the term “Z-Library,” which is
6 undefined. In the context of this Request, Meta will construe this phrase to mean the organization
7 referenced here: <https://z-lib.io/>.

8 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
9 needs of the case to the extent that it is interpreted to seek all documents and communications
10 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
11 and disproportionate to the needs of the case to the extent it seeks documents and communications
12 in the possession, custody, or control of, and/or concerns a third-party.

13 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
14 Protective Order and ESI Order, following the identification of appropriate search terms and
15 custodians by Plaintiffs pursuant to the ESI Order to be used to search for email and electronic
16 communications, and subject to the foregoing objections as well as Meta’s objections to said
17 custodians and search terms, Meta will search for and produce non-privileged, relevant
18 communications responsive to this Request, if any, in accordance with the limitations set forth in
19 the ESI Order. Any such production shall be on a rolling basis, and the timing of any such
20 production shall be governed by terms of the Scheduling Order (ECF No. 87).

21 **REQUEST FOR PRODUCTION NO. 9:**

22 Documents and Communications to, from, or with Sci-Hub Concerning Training Data.

23 **RESPONSE TO REQUEST NO. 9:**

24 Meta incorporates by reference its objections and definitions above, including to the terms
25 “Training Data” and “Communications.” Any electronic communications, including email and
26 documents attached thereto, must be requested pursuant to the ESI Order.

27 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
28 Communications to, from, or with” to the extent it is seeking materials other than those governed

1 by the procedure for requesting email and electronic communications under the ESI Order, and
2 Meta interprets the Request as seeking such email and electronic communications and Documents
3 attached thereto. This Request is also vague and ambiguous as to the term “Sci-Hub,” which is
4 undefined. In the context of this Request, Meta will construe this phrase to mean the organization
5 referenced here: <https://sci-hub.se/database>.

6 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
7 needs of the case to the extent that it is interpreted to seek all documents and communications
8 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
9 and disproportionate to the needs of the case to the extent it seeks documents and communications
10 in the possession, custody, or control of, and/or concerns a third-party.

11 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
12 Protective Order and ESI Order, following the identification of appropriate search terms and
13 custodians by Plaintiffs pursuant to the ESI Order to be used to search for email and electronic
14 communications, and subject to the foregoing objections as well as Meta’s objections to said
15 custodians and search terms, Meta will search for and produce non-privileged, relevant
16 communications responsive to this Request, if any, in accordance with the limitations set forth in
17 the ESI Order. Any such production shall be on a rolling basis, and the timing of any such
18 production shall be governed by terms of the Scheduling Order (ECF No. 87).

19 **REQUEST FOR PRODUCTION NO. 10:**

20 Documents and Communications to, from, or with Bibliotik Concerning Training Data.

21 **RESPONSE TO REQUEST NO. 10:**

22 Meta incorporates by reference its objections and definitions above, including to the terms
23 “Training Data” and “Communications.” Any electronic communications, including email and
24 documents attached thereto, must be requested pursuant to the ESI Order.

25 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
26 Communications to, from, or with” to the extent it is seeking materials other than those governed
27 by the procedure for requesting email and electronic communications under the ESI Order, and
28 Meta interprets the Request as seeking such email and electronic communications and Documents

1 attached thereto. This Request is also vague and ambiguous as to the term “Bibliotik,” which is
2 undefined. Furthermore, the Request is vague, ambiguous, and unintelligible as to “Documents
3 and Communications to, from, or with Bibliotik,” which Meta understands to be the name of a
4 dataset rather than an individual, group, business, or organization.

5 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
6 needs of the case to the extent that it is interpreted to seek all documents and communications
7 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
8 and disproportionate to the needs of the case to the extent it seeks documents and communications
9 in the possession, custody, or control of, and/or concerns what Plaintiffs appear to believe is third-
10 party.

11 **REQUEST FOR PRODUCTION NO. 11:**

12 Documents and Communications to, from, or with The Eye Concerning Training Data.

13 **RESPONSE TO REQUEST NO. 11:**

14 Meta incorporates by reference its objections and definitions above, including to the terms
15 “Training Data” and “Communications.” Any electronic communications, including email and
16 documents attached thereto, must be requested pursuant to the ESI Order.

17 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
18 Communications to, from, or with” to the extent it is seeking materials other than those governed
19 by the procedure for requesting email and electronic communications under the ESI Order, and
20 Meta interprets the Request as seeking such email and electronic communications and Documents
21 attached thereto. This Request is also vague and ambiguous as to the term “The Eye,” which is
22 undefined. In the context of this Request, Meta will construe this phrase to mean the organization
23 referenced here: <https://the-eye.eu/>.

24 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
25 needs of the case to the extent that it is interpreted to seek all documents and communications
26 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
27 and disproportionate to the needs of the case to the extent it seeks documents and communications
28 in the possession, custody, or control of, and/or concerns a third-party.

1 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
2 Protective Order and ESI Order, following the identification of appropriate search terms and
3 custodians by Plaintiffs pursuant to the ESI Order to be used to search for email and electronic
4 communications, and subject to the foregoing objections as well as Meta’s objections to said
5 custodians and search terms, Meta will search for and produce non-privileged, relevant
6 communications responsive to this Request, if any, in accordance with the limitations set forth in
7 the ESI Order. Any such production shall be on a rolling basis, and the timing of any such
8 production shall be governed by terms of the Scheduling Order (ECF No. 87).

9 **REQUEST FOR PRODUCTION NO. 12:**

10 Documents and Communications to, from, or with Anna’s Archive Concerning Training
11 Data.

12 **RESPONSE TO REQUEST NO. 12:**

13 Meta incorporates by reference its objections and definitions above, including to the terms
14 “Training Data” and “Communications.” Any electronic communications, including email and
15 documents attached thereto, must be requested pursuant to the ESI Order.

16 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
17 Communications to, from, or with” to the extent it is seeking materials other than those governed
18 by the procedure for requesting email and electronic communications under the ESI Order, and
19 Meta interprets the Request as seeking such email and electronic communications and Documents
20 attached thereto. This Request is also vague and ambiguous as to the term “Anna’s Archive,” which
21 is undefined. In the context of this Request, Meta will construe this phrase to mean the organization
22 referenced here: <https://annas-archive.org/>.

23 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
24 needs of the case to the extent that it is interpreted to seek all documents and communications
25 concerning the subject matter of the Request. This Request is also overbroad, unduly burdensome,
26 and disproportionate to the needs of the case to the extent it seeks documents and communications
27 in the possession, custody, or control of, and/or concerns a third-party.

28

1 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
2 Protective Order and ESI Order, following the identification of appropriate search terms and
3 custodians by Plaintiffs pursuant to the ESI Order to be used to search for email and electronic
4 communications, and subject to the foregoing objections as well as Meta’s objections to said
5 custodians and search terms, Meta will search for and produce non-privileged, relevant
6 communications responsive to this Request, if any, in accordance with the limitations set forth in
7 the ESI Order. Any such production shall be on a rolling basis, and the timing of any such
8 production shall be governed by terms of the Scheduling Order (ECF No. 87).

9 **REQUEST FOR PRODUCTION NO. 13:**

10 Documents and Communications to, from, or with Hugo Touvron Concerning Training
11 Data.

12 **RESPONSE TO REQUEST NO. 13:**

13 Meta incorporates by reference its objections and definitions above, including to the terms
14 “Training Data” and “Communications.” Any electronic communications, including email and
15 documents attached thereto, must be requested pursuant to the ESI Order.

16 Meta objects to this Request as vague, ambiguous, and unintelligible as to “Documents and
17 Communications to, from, or with” to the extent it is seeking materials other than those governed
18 by the procedure for requesting email and electronic communications under the ESI Order, and
19 Meta interprets the Request as seeking such email and electronic communications and Documents
20 attached thereto.

21 Meta objects to this Request because, on its face, it does not exclude documents and
22 communications exchanged with or at the direction of Meta’s attorneys concerning legal advice or
23 opinions, which are subject to attorney-client privilege and/or attorney work product doctrine. Such
24 documents will not be produced.

25 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
26 needs of the case to the extent that it is interpreted to seek all documents and communications
27 concerning the subject matter of the Request, including documents with limited, if any, relevance
28 to Plaintiffs’ copyright infringement allegations and Meta’s defenses thereto.

1 COOLEY LLP
 BOBBY GHAJAR (198719)
 2 (bghajar@cooley.com)
 COLETTE GHAZARIAN (322235)
 3 (cghazarian@cooley.com)
 1333 2nd Street, Suite 400
 4 Santa Monica, California 90401
 Telephone: (310) 883-6400
 5 MARK WEINSTEIN (193043)
 (mweinstein@cooley.com)
 6 KATHLEEN HARTNETT (314267)
 (khartnett@cooley.com)
 7 JUDD LAUTER (290945)
 (jlauter@cooley.com)
 8 ELIZABETH L. STAMESHKIN (260865)
 (lstameshkin@cooley.com)
 9 3175 Hanover Street
 Palo Alto, CA 94304-1130
 10 Telephone: (650) 843-5000

11 CLEARY GOTTLIEB STEEN & HAMILTON LLP
 ANGELA L. DUNNING (212047)
 12 (adunning@cgsh.com)
 1841 Page Mill Road, Suite 250
 13 Palo Alto, CA 94304
 Telephone: (650) 815-4131

14 *[Full Listing on Signature Page]*
 15 *Counsel for Defendant Meta Platforms, Inc.*

16 **UNITED STATES DISTRICT COURT**
 17 **NORTHERN DISTRICT OF CALIFORNIA**
 18 **SAN FRANCISCO DIVISION**

19 RICHARD KADREY, *et al.*,
 20 Individual and Representative
 21 Plaintiffs,
 22
 23 v.
 24 META PLATFORMS, INC., a Delaware
 corporation;
 25 Defendant.

Case No. 3:23-cv-03417-VC

**DEFENDANT META PLATFORMS, INC.’S
 OBJECTIONS AND RESPONSES TO
 PLAINTIFFS’ FIFTH SET OF REQUESTS FOR
 PRODUCTION**

1 Meta objects to this Request to the extent it seeks information that is not relevant to any
2 party's claims or defenses.

3 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
4 Protective Order and ESI Order, Meta will conduct a reasonable search for and produce non-
5 privileged, responsive documents in its possession, custody, or control, if any, concerning Hugging
6 Face's involvement, if any, in the dissemination of Books3.

7 **REQUEST FOR PRODUCTION NO. 85:**

8 All Documents and Communications related to the decision to use Torrent Systems to
9 acquire data for training Llama Models.

10 **RESPONSE TO REQUEST NO. 85:**

11 Meta incorporates by reference its objections and definitions above. Any electronic
12 communications, including emails and documents attached thereto, will only be produced pursuant
13 to and in accordance with the ESI Order.

14 Meta objects to this Request because, on its face, it does not exclude documents and
15 communications exchanged with or at the direction of Meta's attorneys concerning legal advice or
16 opinions, which are subject to attorney-client privilege and/or attorney work product
17 doctrine. Such documents will not be produced.

18 Meta objects to this Request as vague and ambiguous as to the terms "decision to use" and
19 "acquire data." Meta construes this Request as seeking documents related to any decision by Meta
20 to use data accessed through Torrent Systems (as construed above) for training Llama Models (as
21 construed above).

22 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
23 needs of the case to the extent that it seeks all documents and communications concerning the
24 subject matter of the Request, including documents and communications with limited, if any,
25 relevance to Plaintiffs' copyright infringement allegations and Meta's defenses thereto.

26 Meta objects to this Request to the extent it seeks information that is not relevant to any
27 party's claims or defenses.

28 Subject to and without waiving the foregoing objections, and pursuant to the terms of the

1 Protective Order and ESI Order, Meta will conduct a reasonable search for and produce non-
2 privileged, responsive documents in its possession, custody, or control, if any, concerning any
3 decision by Meta to use Torrent Systems (as construed above) to acquire data for training the Llama
4 Models (as construed above).

5 **REQUEST FOR PRODUCTION NO. 86:**

6 All Documents and Communications related to the decision to use books for long-range
7 context modeling research.

8 **RESPONSE TO REQUEST NO. 86:**

9 Meta incorporates by reference its objections and definitions above. Any electronic
10 communications, including emails and documents attached thereto, will only be produced pursuant
11 to and in accordance with the ESI Order.

12 Meta objects to this Request because, on its face, it does not exclude documents and
13 communications exchanged with or at the direction of Meta’s attorneys concerning legal advice or
14 opinions, which are subject to attorney-client privilege and/or attorney work product
15 doctrine. Such documents will not be produced.

16 Meta objects to this Request on the ground that it assumes that Meta used books for long-
17 range context research.

18 Meta objects to this Request as vague and ambiguous as to the terms “decision,” which is
19 non-specific as to the purported decisionmaker at issue, and “use of books,” which is not limited in
20 time or scope or to particular LLMs. Meta construes this Request as seeking documents concerning
21 Meta’s decision, if any, to use books as training data for long-range context research related to the
22 Llama Models (as construed above).

23 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
24 needs of the case to the extent that it seeks all documents and communications concerning the
25 subject matter of the Request, including documents and communications with limited, if any,
26 relevance to Plaintiffs’ copyright infringement allegations and Meta’s defenses thereto.

27 Meta objects to this Request to the extent it seeks information that is not relevant to any
28 party’s claims or defenses.

1 Meta objects to this Request as vague, ambiguous, and indefinite as to “copyrighted
2 material,” as the Request does not identify any and Meta is not in a position to know whether any
3 particular work is subject to copyright protection. Meta construes this Request as seeking
4 documents concerning Meta’s training data memorization mitigations for the Llama Models (as
5 construed above).

6 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
7 needs of the case to the extent that it seeks all documents and communications, including source
8 code, concerning the subject matter of the Request, including documents and communications with
9 limited, if any, relevance to Plaintiffs’ copyright infringement allegations and Meta’s defenses
10 thereto.

11 Meta objects to this Request to the extent it seeks information that is not relevant to any
12 party’s claims or defenses.

13 Meta objects to this Request as duplicative of other Requests, including Request No. 116.

14 Meta objects to this Request to the extent that Meta has already searched for and produced
15 documents responsive to this Request, which overlaps with prior Requests.

16 Subject to and without waiving the foregoing objections, and pursuant to the terms of the
17 Protective Order and ESI Order, Meta will conduct a reasonable search for and produce non-
18 privileged, responsive documents in its possession, custody, or control, if any, sufficient to show
19 Meta’s training data memorization mitigations for the Llama Models (as construed above).

20 **REQUEST FOR PRODUCTION NO. 119:**

21 All Documents and Communications, including source code, relating to the processing of
22 copyrighted material used in training Llama Models, including storage and deletion of copyrighted
23 material.

24 **RESPONSE TO REQUEST NO. 119:**

25 Meta incorporates by reference its objections and definitions above. Any electronic
26 communications, including emails and documents attached thereto, will only be produced pursuant
27 to and in accordance with the ESI Order.

28 Meta objects to this Request because, on its face, it does not exclude documents and

1 communications exchanged with or at the direction of Meta’s attorneys concerning legal advice or
2 opinions, which are subject to attorney-client privilege and/or attorney work product
3 doctrine. Such documents will not be produced.

4 Meta objects to this Request as vague, ambiguous, and indefinite as to “copyrighted
5 material,” as the Request does not identify any and Meta is not in a position to know whether any
6 particular training data is protected by copyright. Meta also objects to this Request on the ground
7 that it presupposes that there are documents related to the processing, storage, and deletion of
8 copyrighted material separate and apart from other training data.

9 Meta objects to this Request as duplicative of other Requests.

10 Meta objects to this Request as overbroad, unduly burdensome, and disproportionate to the
11 needs of the case to the extent that it seeks all documents and communications concerning the
12 subject matter of the Request, including documents and communications, including source code,
13 with limited, if any, relevance to Plaintiffs’ copyright infringement allegations and Meta’s defenses
14 thereto.

15 Meta objects to this Request to the extent it seeks information that is not relevant to any
16 party’s claims or defenses.

17 Meta objects to this Request as duplicative of other Requests, including Request Nos. 116
18 and 118.

19 Meta is willing to meet and confer with Plaintiffs regarding the appropriate scope, if any of
20 this Request.

21 **REQUEST FOR PRODUCTION NO. 120:**

22 All Documents and Communications, including source code, relating to actual or
23 contemplated source code changes within Llama Models, including source code “commits” and
24 “pull requests.”

25 **RESPONSE TO REQUEST NO. 120:**

26 Meta incorporates by reference its objections and definitions above. Any electronic
27 communications, including emails and documents attached thereto, will only be produced pursuant
28 to and in accordance with the ESI Order.

EXHIBIT B

FILED UNDER SEAL

EXHIBIT C

FILED UNDER SEAL

EXHIBIT D

BOIES SCHILLER FLEXNER LLP

David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsflfp.com

Maxwell V. Pritt (SBN 253155)
Joshua I. Schiller (SBN 330653)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsflfp.com
jischiller@bsflfp.com
jstein@bsflfp.com

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

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

*Counsel for Individual and Representative Plaintiffs
and the Proposed Class. (additional counsel included below)*

**UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION**

Richard Kadrey, et al.,

Individual and Representative Plaintiffs,

v.

Meta Platforms, Inc.,

Defendant.

JOSEPH SAVERI LAW FIRM, LLP

Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Holden Benon (SBN 325847)
Aaron Cera (SBN 351163)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
hbenon@saverilawfirm.com
acera@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@buttericklaw.com

**CAFFERTY CLOBES MERIWETHER &
SPRENGEL LLP**

Bryan L. Clobes (*pro hac vice*)
135 S. LaSalle Street, Suite 3210
Chicago, IL 60603
(312)-782-4880
bclobes@caffertyclobes.com

Case No. 3:23-cv-03417-VC

**PLAINTIFFS' FIFTH SET OF REQUESTS
FOR PRODUCTION TO DEFENDANT
META**

1 **REQUESTS FOR PRODUCTION OF DOCUMENTS**

2 **REQUEST FOR PRODUCTION NO. 79**

3 All Documents and Communications related to the decision to describe the Books3 database as
4 “publicly available” in the Llama 1 Paper.

5 **REQUEST FOR PRODUCTION NO. 80**

6 All Documents and Communications related to the decision to describe the Books3 database as
7 “publicly available” in the Llama 2 Paper.

8 **REQUEST FOR PRODUCTION NO. 81**

9 All Documents and Communications related to the decision to use Shadow Datasets for training
10 Llama Models.

11 **REQUEST FOR PRODUCTION NO. 82**

12 All Documents and Communications related to the role of EleutherAI in the acquisition and use of
13 Books3.

14 **REQUEST FOR PRODUCTION NO. 83**

15 All Documents and Communications related to the role of The Eye in the distribution of Books3.

16 **REQUEST FOR PRODUCTION NO. 84**

17 All Documents and Communications related to the role of Hugging Face in the distribution of
18 Books3.

19 **REQUEST FOR PRODUCTION NO. 85**

20 All Documents and Communications related to the decision to use Torrent Systems to acquire data
21 for training Llama Models.

22 **REQUEST FOR PRODUCTION NO. 86**

23 All Documents and Communications related to the decision to use books for long-range context
24 modeling research.

25 **REQUEST FOR PRODUCTION NO. 87**

26 All Documents and Communications related to the decision to use books for coherent storytelling
27 in the training of Llama Models.

REQUEST FOR PRODUCTION NO. 129

Documents sufficient to identify all individuals, currently or formerly employed by Meta, including their roles and titles, who work or worked on Meta’s AI research and development projects, including Llama models, and/or on Meta’s licensing efforts with respect to AI training data.

REQUEST FOR PRODUCTION NO. 130

All Documents and Communications, including discussions, deliberations, or negotiations related to any actual, proposed, or contemplated licensing agreements (even if never written or executed) for AI training data, including any actual, proposed, or contemplated terms, conditions, and consideration.

Dated: October 9, 2024

By: /s/ David L. Simons

JOSEPH SAVERI LAW FIRM, LLP

Joseph R. Saveri (SBN 130064)
Cadio Zirpoli (SBN 179108)
Christopher K.L. Young (SBN 318371)
Holden Benon (SBN 325847)
Aaron Cera (SBN 351163)
601 California Street, Suite 1505
San Francisco, California 94108
(415) 500-6800
jsaveri@saverilawfirm.com
czirpoli@saverilawfirm.com
cyoung@saverilawfirm.com
hbenon@saverilawfirm.com
acera@saverilawfirm.com

Matthew Butterick (SBN 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
(323) 968-2632
mb@buttericklaw.com

BOIES SCHILLER FLEXNER LLP

David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsflfp.com

Maxwell V. Pritt (SBN 253155)
Joshua I. Schiller (SBN 330653)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsflfp.com
jischiller@bsflfp.com
jstein@bsflfp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005

1 **CAFFERTY CLOBES MERIWETHER &**
2 **SPRENGEL LLP**

3 Bryan L. Clobes (*pro hac vice*)
4 135 S. LaSalle Street, Suite 3210
5 Chicago, IL 60603
6 (312)-782-4880
7 bclobes@caffertyclobes.com

(202) 237-2727
jpanuccio@bsfllp.com

8 David L. Simons (*pro hac vice*)
9 55 Hudson Yards, 20th Floor
10 New York, NY 10001
11 (914) 749-8200
12 dsimons@bsfllp.com

13 **DICELLO LEVITT LLP**

14 Amy Keller (*pro hac vice*)
15 Nada Djordjevic (*pro hac vice*)
16 James Ulwick (*pro hac vice*)
17 (312) 214-7900
18 akeller@dicellolevitt.com
19 ndjordjevic@dicellolevitt.com
20 julwick@dicellolevitt.com
21 10 North Dearborn Street, Sixth Floor
22 Chicago, Illinois 60602

23 David Straite (*pro hac vice*)
24 (646) 933-1000
25 dsraite@dicellolevitt.com
26 485 Lexington Avenue
27 Suite 1001
28 New York, New York 10017

CERTIFICATE OF SERVICE

I, the undersigned, am employed by the Boies Schiller Flexner LLP. My business address is 55 Hudson Yards, New York, NY 10001. I am over the age of eighteen and not a party to this action.

On October 9, 2024, I caused the following documents to be served by email upon the parties listed on the attached Service List:

- **PLAINTIFFS’ FIFTH SET OF REQUESTS FOR PRODUCTION TO DEFENDANT META**

I declare under penalty of perjury that the foregoing is true and correct. Executed on October 9, 2024.

/s/ David L. Simons

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

SERVICE LIST

1
2 Bobby A. Ghajar
3 Colette Ani Ghazarian
4 **COOLEY LLP**
5 1333 2nd Street, Suite 400
6 Santa Monica, CA 90401
7 Email: bghajar@cooley.com
8 cghazarian@cooley.com

9 Kathleen R. Hartnett
10 **COOLEY LLP**
11 3 Embarcadero Center, 20th Floor
12 San Francisco, CA 94111-4004
13 Email: khartnett@cooley.com

14 Judd D. Lauter
15 **COOLEY LLP**
16 3175 Hanover Street
17 Palo Alto, CA 94304
18 Email: jlauter@cooley.com

19 Mark Alan Lemley
20 **LEX LUMINA PLLC**
21 745 Fifth Avenue, Suite 500
22 New York, NY 10151
23 Email: mlemley@lex-lumina.com

24 Angela L. Dunning
25 **CLEARY GOTTLIEB STEEN &**
26 **HAMILTON LLP**
27 1841 Page Mill Road
28 Palo Alto, CA 94304-1254
Email: adunning@cgsh.com

*Counsel for Defendant
Meta Platforms, Inc.*

EXHIBIT E

1 Joseph R. Saveri (State Bar No. 130064)
 Cadio Zirpoli (State Bar No. 179108)
 2 Christopher K.L. Young (State Bar No. 318371)
 Holden Benon (State Bar No. 325847)
 3 Kathleen J. McMahon (State Bar No. 340007)
 4 **JOSEPH SAVERI LAW FIRM, LLP**
 601 California Street, Suite 1000
 5 San Francisco, CA 94108
 Telephone: (415) 500-6800
 6 Facsimile: (415) 395-9940
 Email: jsaveri@saverilawfirm.com
 7 czirpoli@saverilawfirm.com
 8 cyoung@saverilawfirm.com
 hbenon@saverilawfirm.com
 9 kmcmahon@saverilawfirm.com

10 Matthew Butterick (State Bar No. 250953)
 11 1920 Hillhurst Avenue, #406
 Los Angeles, CA 90027
 12 Telephone: (323) 968-2632
 Facsimile: (415) 395-9940
 13 Email: mb@buttericklaw.com

14 *Counsel for Individual and Representative Plaintiffs*
 15 *and the Proposed Class*

16 [Additional Counsel Listed on Signature Page]

17 **UNITED STATES DISTRICT COURT**
 18 **NORTHERN DISTRICT OF CALIFORNIA**
 19 **SAN FRANCISCO DIVISION**

20 Richard Kadrey, Sarah Silverman, Christopher
 Golden, Michael Chabon, Ta-Nehisi Coates, Junot
 21 Díaz, Andrew Sean Greer, David Henry Hwang,
 Matthew Klam, Laura Lippman, Rachel Louise
 22 Snyder, Ayelet Waldman, and Jacqueline Woodson,

23 *Individual and Representative Plaintiffs,*

24 v.

25 Meta Platforms, Inc., a Delaware corporation;

26 *Defendant.*

Case No. 3:23-cv-03417-VC

**PLAINTIFFS' FIRST SET OF
 REQUESTS FOR PRODUCTION TO
 DEFENDANT META**

1 8. In responding to these Requests for Production, You are to Include Documents (1)
2 obtained from witnesses who gave information to any governmental agency or investigatory body,
3 including Congress; (2) that constitute, refer or relate to summaries of testimony or other statements in
4 connection with any governmental agency or investigatory body proceeding or investigation, including
5 before Congress; or (3) obtained on Your behalf in preparation for testimony or interviews before any
6 governmental agency or investigatory body, including Congress.

7
8 **REQUESTS FOR PRODUCTION OF DOCUMENTS**

9 **REQUEST FOR PRODUCTION NO. 1**

10 The Training Data for Llama 1.

11 **REQUEST FOR PRODUCTION NO. 2**

12 The Training Data for Llama 2.

13 **REQUEST FOR PRODUCTION NO. 3**

14 The Training Data for Llama 3.

15 **REQUEST FOR PRODUCTION NO. 4**

16 The Training Data comprising the “Books3” dataset referenced in the paper “LLaMA: Open and
17 Efficient Foundation Language Models,” and all Documents and Communications Concerning Your
18 selection and assembly of this dataset.

19 **REQUEST FOR PRODUCTION NO. 5**

20 The Training Data comprising the “Gutenberg” dataset referenced in the paper “LLaMA: Open
21 and Efficient Foundation Language Models”, and all Documents and Communications Concerning
22 Your selection and assembly of this dataset.

23 **REQUEST FOR PRODUCTION NO. 6**

24 Documents and Communications to, from, or with Project Gutenberg Concerning Training
25 Data.

26 **REQUEST FOR PRODUCTION NO. 7**

27 Documents and Communications to, from, or with Library Genesis (aka LibGen) Concerning
28 Training Data.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

Bryan L. Clobes (*pro hac vice anticipated*)
**CAFFERTY CLOBES MERIWETHER
& SPRENGEL LLP**
205 N. Monroe Street
Media, PA 19063
Tel: 215-864-2800
bclobes@caffertyclobes.com

Alexander J. Sweatman (*pro hac vice anticipated*)
**CAFFERTY CLOBES MERIWETHER
& SPRENGEL LLP**
135 South LaSalle Street, Suite 3210
Chicago, IL 60603
Tel: 312-782-4880
asweatman@caffertyclobes.com

*Counsel for Individual and Representative
Plaintiffs and the Proposed Class*

PROOF OF SERVICE

I, the undersigned, am employed by the Joseph Saveri Law Firm, LLP. My business address is 601 California Street, Suite 1000, San Francisco, California 94108. I am over the age of eighteen and not a party to this action.

On December 27, 2023, I caused the following documents to be served by email upon all persons appearing on the attached Service List:

- **PLAINTIFFS' FIRST SET OF REQUESTS FOR PRODUCTION TO DEFENDANT META**

I declare under penalty of perjury that the foregoing is true and correct. Executed on December 27, 2023.



Ruby Ponce

SERVICE LIST

1
2 Bobby A. Ghajar
3 Colette Ani Ghazarian
4 **COOLEY LLP**
5 1333 2nd Street, Suite 400
6 Santa Monica, CA 90401
7 Email: bghajar@cooley.com
8 cghazarian@cooley.com

Judd D. Lauter
COOLEY LLP
3175 Hanover Street
Palo Alto, CA 94304
Email: jlauter@cooley.com

9 Kathleen R. Hartnett
10 **COOLEY LLP**
11 3 Embarcadero Center, 20th Floor
12 San Francisco, CA 94111-4004
13 Email: khartnett@cooley.com

Mark Alan Lemley
LEX LUMINA PLLC
745 Fifth Avenue, Suite 500
New York, NY 10151
Email: mlemley@lex-lumina.com

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28