

EXHIBIT H

DUNN ISAACSON RHEE LLP
KAREN L. DUNN (*pro hac vice*)
kdunn@dirllp.com
JESSICA E. PHILLIPS (*pro hac vice*)
jphillips@dirllp.com
KYLE N. SMITH (*pro hac vice*)
ksmith@dirllp.com
401 9th Street NW
Washington, DC 20004
Telephone: (202) 240-2900

CLEARY GOTTlieb STEEN &
HAMILTON LLP
ANGELA L. DUNNING (212047)
adunning@cgsh.com
151 University Avenue
Palo Alto, CA 94301-1632
Telephone: (650) 815-4131

Counsel for Defendant Meta Platforms, Inc.

COOLEY LLP
MARK R. WEINSTEIN (193043)
mweinstein@cooley.com
3175 Hanover Street
Palo Alto, CA 94304-1130
Telephone: (650) 843-5000

PHILLIP E. MORTON (*pro hac vice*)
pmorton@cooley.com
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004
Telephone: (202) 842-7800

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

CAMBRONNE, INC., et al.

Plaintiff,

v.

META PLATFORMS, INC.,

Defendant.

Case No. 3:26-cv-03725-VC

**DEFENDANT META PLATFORMS, INC.'S 1ST
SUPPLEMENTAL RESPONSES AND
OBJECTIONS TO PLAINTIFF'S FIRST SET OF
INTERROGATORIES**

Judge: Hon. Vince Chhabria

PROPOUNDING PARTY: PLAINTIFFS CAMBRONNE INC., LISA BARRETTA, PHILIP SHISHKIN, JANE ADAMS, MATTHEW SACKS, AND MICHAEL KOCHIN

RESPONDING PARTY: DEFENDANT META PLATFORMS, INC.

SET NUMBER: ONE

Pursuant to Federal Rules of Civil Procedure Rules 26 and 33, and Local Rule 33 of the United States District Court for the Northern District of California, Defendant Meta Platforms, Inc. (“Meta”) hereby objects and responds as follows to Plaintiffs Cambronne Inc., Lisa Barretta, Philip Shishkin, Jane Adams, Matthew Sacks, and Michael Kochin’s (“Plaintiffs”) First Set of Interrogatories (“Interrogatories”).

I. Responses to All Interrogatories

1. Meta’s responses to these Interrogatories are made to the best of Meta’s present knowledge, information, and belief. Said responses are at all times subject to such additional or different information that discovery or further investigation may disclose, and while based on the present state of Meta’s recollection, are subject to such refreshing of recollection, and such additional knowledge of facts, as may result from Meta’s further discovery or investigation. Meta reserves the right to make any use of, or to introduce at any hearing and at trial, information and/or documents responsive to these Interrogatories but discovered subsequent to the date of these responses, including, but not limited to, any such information or documents obtained in discovery herein.

2. To the extent that Meta responds to an Interrogatory by stating that Meta will provide information or documents that Meta deems to embody material that is private, business confidential, proprietary, trade secret, or otherwise protected from disclosure pursuant to Federal Rule of Civil Procedure 26(c) and/or Federal Rule of Evidence 501, Meta will only do so subject to the Protective Order (Dkt. 53) governing the unauthorized use or disclosure of such information or documents with a designated “CONFIDENTIAL” or “HIGHLY CONFIDENTIAL - ATTORNEYS’ EYES ONLY” or “HIGHLY CONFIDENTIAL – SOURCE CODE” distinction.

3. Meta reserves all objections or other questions as to the competency, relevance, materiality, privilege or admissibility as evidence in any subsequent proceeding in or trial of this or any other action for any purpose whatsoever of Meta’s responses herein and any document or thing identified

1 or provided in response to the Interrogatories.

2 **II. Objections to Instructions and Definitions**

3 Whether or not separately set forth in response to each Interrogatory, Meta makes these
4 objections to the following Instructions and Definitions.

5 1. Meta objects to the definition of “COPYRIGHT MANAGEMENT
6 INFORMATION” to the extent it calls for a legal conclusion. Specifically, a determination of
7 whether any particular data, text, or metadata constitutes “Copyright Management Information”
8 under 17 U.S.C. § 1202(c) is inherently legal in nature and calls for a legal conclusion. Meta is not
9 in a position to know whether any particular information qualifies as legally protected Copyright
10 Management Information and a determination of such is not within Meta’s possession, custody, or
11 control. Meta further objects to this definition as vague and ambiguous, overbroad, unduly
12 burdensome, and disproportionate to the needs of the case to the extent it includes identifying
13 information for persons that “created” copyrighted works.

14 2. Meta objects to the definition of “PLAINTIFFS’ WORK(S)” as overbroad, unduly
15 burdensome and disproportionate to the needs of the case to the extent it puts the burden on Meta
16 to determine which works purportedly “belong” to PLAINTIFFS. Meta further objects to the term
17 “belong” as vague and ambiguous as it is unclear what form of possession or ownership is intended
18 by the use of this term. Without agreeing that PLAINTIFFS possesses or owns any valid copyright
19 in any specific works, Meta will interpret the term PLAINTIFFS’ WORKS to refer to the specific
20 works identified in Exhibit A to the Interrogatories.

21 3. Meta objects to the definitions of “COMMUNICATIONS” and “DOCUMENT(S)”
22 to the extent these definitions are inconsistent with and otherwise seek to circumvent the custodian
23 and search term limits for electronic communications (including emails and other electronic
24 correspondence, and documents attached thereto), as provided in the forthcoming ESI Order.

25 4. Meta further objects to the definitions of “COMMUNICATIONS” and
26 “DOCUMENT(S)” to the extent they seek to expand or alter the scope of the Federal Rules of Civil
27 Procedure, the Local Rules of this Court, or the forthcoming ESI Order. Meta further objects to
28 these definitions to the extent they seek Meta to produce documents that are not reasonably

1 accessible because of undue burden or cost pursuant to Fed. R. Civ. P. 26(b)(2)(B), which are
2 unduly burdensome to preserve and search, or which otherwise, under the proportionality factors,
3 should not be preserved.

4 **5.** Meta objects to the definitions of “CORPORA” and “DATASET(S)” as overbroad,
5 unduly burdensome, and disproportionate to the needs of the case, including (but not limited to) the
6 extent they refer to non-textual materials that are not relevant to any party’s claims or defenses and
7 to the extent they refer vaguely to “other content.” Without any admission with respect to content
8 or nature, Meta will construe the term “DATASET(S)” to refer to Anna’s Archive, Bibliotik,
9 Books3, C4, Library Genesis (or LibGen), Sci-Hub, The Pile, and Z-Library.

10 **6.** Meta objects to the definition of “INFRINGE,” “INFRINGED,” and “INFRINGES”
11 to the extent it calls for a legal conclusion. Specifically, a determination of whether any conduct
12 constitutes copyright infringement is inherently legal in nature and calls for a legal conclusion.
13 Meta is not in a position to know whether any particular conduct infringes on a legally protected
14 work and a determination of such is not within Meta’s possession, custody, or control.

15 **7.** Meta objects to the definition of “LARGE LANGUAGE MODEL(S)” and “LLM
16 MODEL(S)” as vague and ambiguous, overbroad, unduly burdensome, and disproportionate to the
17 needs of the case. In particular, the phrase “any machine learning model, neural network,
18 generative artificial intelligence model, transformer model, or similar computational system trained
19 on large quantities of text or other data” is overly broad and encompasses tools and services that
20 are not relevant to any party’s claims or defenses. Meta further objects to this definition to the
21 extent that it purports to require Meta to produce documents or information concerning large
22 language models that were not publicly released and/or were not trained on corpuses of text that
23 allegedly include any of Plaintiff’s allegedly copyrighted works. Meta further objects that the
24 phrase “process, predict, generate, summarize, classify, or otherwise manipulate natural language
25 or content” renders the definition vague, ambiguous, confusing, and overly broad. Meta will
26 construe “LARGE LANGUAGE MODEL(S)” and “LLM MODEL(S)” to mean the publicly
27 released Llama 1, Llama 2, Llama 3, Llama 4, and Muse Spark models.

28 **8.** Meta objects to the definition of “META,” “YOU,” and “YOUR” as overbroad,

1 unduly burdensome, and disproportionate to the needs of the case. Meta will construe “META,”
2 “YOU,” or “YOUR” to mean Meta Platforms, Inc.

3 **9.** Meta objects to the definition of “PIRATED,” “PIRACY,” and “PIRATED
4 MATERIAL” to the extent it calls for legal conclusions and assumes disputed issues. Meta further
5 objects that the definition is phrased in a way that is argumentative, prejudicial, and lacks
6 foundation. Meta is not in a position to know whether any particular work is subject to copyright
7 protection or whether third-party repositories lacked authorization to host or possess such works.

8 **10.** Meta objects to the definition of “SHADOW LIBRARY” and “SHADOW
9 LIBRARIES” to the extent it calls for legal conclusions and assumes disputed issues. Meta further
10 objects that the definition is phrased in a way that is argumentative, prejudicial, and lacks
11 foundation. Meta further objects to this definition as overbroad, unduly burdensome, and
12 disproportionate to the needs of the case, including (but not limited to) the extent it refers to non-
13 textual materials that are not relevant to any party’s claims or defenses and to the extent it refers
14 vaguely to “media and similar materials.” Meta is not in a position to know whether any particular
15 work is subject to copyright protection or whether third-party repositories lacked authorization to
16 host or possess such works. Meta further objects to this definition to the extent that it incorporates
17 the objectionable portions of the term “DATASETS.” Without any admission with respect to
18 content or nature, Meta will construe the term “SHADOW LIBRARY” to refer to certain of the
19 third-party datasets identified by Plaintiff, namely Anna’s Archive, Bibliotik, Books3, Library
20 Genesis (or LibGen), Sci-Hub, The Pile, and Z-Library.

21 **11.** Meta objects to the definition of “WEB SCRAPING” as vague, ambiguous,
22 overbroad, unduly burdensome, and disproportionate to the needs of the case. In particular, the
23 extensive list of undefined terms, including “extraction, collection, copying, downloading,
24 crawling, indexing, harvesting, or retrieval” and “aggregation, storage, processing” are susceptible
25 to multiple definitions, rendering this definition unintelligibly ambiguous. Meta further objects to
26 the phrase “images, metadata, or other content” as overbroad and disproportionate to the needs of
27 the case, to the extent it refers to training data other than text data that is not relevant to any party’s
28 claims or defenses. Meta further objects to this definition to the extent that it incorporates the

1 objectionable portions of the terms “SHADOW LIBRARIES” and “LARGE LANGUAGE
2 MODELS.” Meta will construe “WEB SCRAPING” to mean the automated collection of text data
3 from the public internet for use in training the Meta Language Models (as construed above).

4 **III. General Objections**

5 Meta hereby incorporates the following general objections into each of its specific
6 responses and objections to the individual Interrogatories as if fully set forth therein:

7 1. Meta incorporates by reference each and every General Objection set forth below
8 into each and every specific Response. From time to time, a specific Response may repeat a
9 General Objection for emphasis or some other reason. The failure to repeat any General Objection
10 in any specific Response shall not be interpreted as a waiver of any General Objection to that
11 Response.

12 2. Any information provided in response to any individual Interrogatory is made
13 without waiver of, and subject to, the General Objections and any additional objections asserted in
14 response to that specific Interrogatory.

15 3. Meta objects to the Interrogatories to the extent that they impose any differing or
16 additional obligations from those required by the Federal Rules, the Local Rules, or any other
17 applicable rule, law, or doctrine, the forthcoming ESI Order and the Protective Order entered in
18 this action (the “Applicable Laws, Rules, and Agreements”), on the ground that such a requirement
19 exceeds the scope of permissible discovery and attempts to impose an unreasonable burden and
20 expense on Meta. To the extent any Interrogatory conflicts with, or purports to impose additional
21 or different duties and obligations, Meta will respond in accordance with the Applicable Laws,
22 Rules, and Agreements.

23 4. Meta objects to the Interrogatories to the extent that they are overly broad, vague,
24 ambiguous, and unduly burdensome; or seek documents and information that are neither relevant
25 nor proportional to the needs of this action.

26 5. Meta objects to the Interrogatories to the extent they seek information or documents
27 or seek to impose a search for information or documents: (a) not within Meta’s possession, custody,
28 or control; (b) not maintained by Meta in the normal course of business; (c) already in the

1 possession, custody, or control of Plaintiff; (d) readily available through public sources or from
2 sources that are more convenient, less burdensome, or less expensive; or (e) from sources that are
3 more readily available to Plaintiff than to Meta. Meta further objects to the extent the burden or
4 expense of the proposed discovery outweighs its likely benefit.

5 **6.** Meta objects to the Interrogatories to the extent that they seek documents and
6 information protected from disclosure by privileges and other protections from production,
7 including, without limitation: (a) the attorney-client privilege; (b) the work-product doctrine; or (c)
8 any other constitutional, statutory, common law or regulatory protection, immunity, or proscription
9 from disclosure. In responding to the Interrogatories, Meta will not produce information or
10 documents protected by such privileges. Where Meta withholds information under claim of
11 privilege, Meta will adhere to the requirements under the forthcoming ESI Order. Meta does not
12 intend the inadvertent production of any privileged or protected information to constitute a waiver
13 of Meta's rights to assert any applicable privilege or protection, and Meta further reserves the right
14 to demand that Plaintiff return, destroy, or sequester any privileged or protected documents
15 inadvertently produced and all copies thereof consistent with Fed. R. Civ. P. 26(b)(5)(B) and the
16 terms and conditions of the forthcoming ESI Order and the Protective Order entered in this action.

17 **7.** Meta objects to the Interrogatories to the extent that they purport to require Meta to
18 provide information or documents in violation of a legal, contractual, or other obligation of
19 nondisclosure owed to a third party. Meta will not provide information or documents subject to
20 any such obligation unless and until it receives the necessary consent from any such third party or,
21 in the absence of such consent, unless and until it is ordered to do so by a court of competent
22 jurisdiction.

23 **8.** Meta objects to the Interrogatories to the extent they call for, or can be interpreted
24 as calling for, Meta to make legal conclusions, or to the extent they assume disputed issues or are
25 phrased in such a way as to be lacking foundation, argumentative, prejudicial, or otherwise
26 improper.

27 **9.** Meta objects to the Interrogatories to the extent that they seek confidential, trade
28 secret, competitively sensitive, business, financial, or other proprietary information. To the extent

documents or information containing any such information are nonprivileged, responsive to the Interrogatories, and are not otherwise objectionable, Meta will only produce such documents and information subject to the forthcoming ESI Order and the Protective Order entered in this action. Meta further reserves the right to redact or otherwise withhold any confidential, proprietary, trade secret, or competitively sensitive information that is neither relevant nor proportional to the needs of this action.

10. Meta objects to the Interrogatories to the extent they seek documents that are not proportionate to the needs of the case and/or not relevant to any party's claims or defenses.

11. Meta objects to the Interrogatories that seek to impose an obligation to investigate or discover information or materials from third parties or sources that are equally accessible to Plaintiff as duplicative, overbroad, unduly burdensome, and oppressive.

12. Meta reserves the right to assert additional objections, or to supplement its Responses as appropriate, particularly if any additional information regarding the Interrogatories or the underlying claims at issue is provided.

IV. OBJECTIONS AND RESPONSES TO INDIVIDUAL INTERROGATORIES

INTERROGATORY NO. 1:

Which PLAINTIFFS' WORKS did YOU obtain from a SHADOW LIBRARY?

RESPONSE TO INTERROGATORY NO. 1:

Meta incorporates by reference its objections and definitions above, including to the terms PLAINTIFFS' WORKS, YOU, and SHADOW LIBRARY.

Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable investigation to identify which PLAINTIFFS' WORKS, if any, Meta downloaded as part of Anna's Archive, Books3 and LibGen for use in training Meta's LARGE LANGUAGE MODELS. As previously communicated to counsel for PLAINTIFFS, Meta may need additional information, including the text of the PLAINTIFFS' WORKS, in order to provide a more complete response to this Interrogatory. Meta will supplement this response on a rolling basis as it identifies responsive information.

SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 1 (AUGUST 24, 2026):

Subject to and without waiving the foregoing objections, and following a reasonable inquiry, Meta further responds as follows:

Attached as **Exhibit A** hereto is a table identifying, based on Meta's current knowledge, information, and belief based on a reasonable investigation, the authors and titles of PLAINTIFFS' WORKS that Meta believes were downloaded from Libgen, Zlib, and Books3.

Based on Meta's current knowledge, information and belief based on a reasonable investigation, to date Meta has not identified PLAINTIFFS' WORKS works that were downloaded from the Internet Archive Lending Library.

Meta's investigation is ongoing and Meta will supplement this Interrogatory as additional responsive information becomes available.

INTERROGATORY NO. 2:

Identify each PLAINTIFFS' WORK that YOU obtained through WEB SCRAPING.

RESPONSE TO INTERROGATORY NO. 2:

Meta incorporates by reference its objections and definitions above, including to the terms PLAINTIFFS' WORK, YOU, and WEB SCRAPING.

Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable investigation to identify which PLAINTIFFS' WORKS, if any, Meta downloaded as part of any web archiving activity conducted by Meta's SpiderMate web archiver for use in training Meta's LARGE LANGUAGE MODELS. As previously communicated to counsel for PLAINTIFFS, Meta may need additional information, including the text of the PLAINTIFFS' WORKS, in order to provide a more complete response to this Interrogatory. Meta will supplement this response on a rolling basis as it identifies responsive information.

SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 2 (AUGUST 24, 2026):

Subject to and without waiving the foregoing objections, and following a reasonable inquiry, Meta further responds as follows:

Based on Meta's current knowledge, information, and belief based on reasonable investigation, Meta has not identified any PLAINTIFFS' WORKS that Meta believes were obtained from Common Crawl, an open repository of web crawl data made available by the Common Crawl Foundation.

Meta also operates a web crawler "SpiderMate" that obtains publicly available web data from the Internet. The SpiderMate data is organized and searchable based on web domains (e.g. domain portion of Uniform Resource Locators (URLs)). Meta is willing to search SpiderMate data for a reasonable number of web domains identified by PLAINTIFFS upon request.

Meta's investigation is ongoing and Meta will supplement this Interrogatory as additional responsive information becomes available.

INTERROGATORY NO. 3:

For each PLAINTIFFS' WORK that YOU obtained from a SHADOW LIBRARY or through WEB SCRAPING, when and how did you obtain it? Identify and describe each instance in detail, including the source or sources, date or dates of acquisition, and means or method of acquisition.

RESPONSE TO INTERROGATORY NO. 3:

Meta incorporates by reference its objections and definitions above, including to the terms PLAINTIFFS' WORK, YOU, SHADOW LIBRARY, and WEB SCRAPING.

Meta objects to this request to the extent it is duplicative of information sought by Interrogatory Nos. 1 and 2. Meta further objects to this interrogatory as vague, ambiguous, overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent the phrase "[i]dentify and describe each instance in detail, including" seeks unspecified information

beyond the recited request for “the source or sources, date or dates of acquisition, and means or method of acquisition” of data obtained from a SHADOW LIBRARY or WEB SCRAPING.

Subject to and without waiving the foregoing objections, as stated for Interrogatory Nos. 1 and 2, Meta is undertaking a reasonable investigation to identify which PLAINTIFFS’ WORKS, if any (1) Meta downloaded as part of Anna’s Archive, Books3, and/or LibGen for use in training Meta’s LARGE LANGUAGE MODELS and (2) Meta downloaded as part of any web archiving activity conducted by Meta’s SpiderMate web archiver for use in training Meta’s LARGE LANGUAGE MODELS. Meta will provide reasonably available information regarding the source, timing, and method of download. As previously communicated to counsel for PLAINTIFFS, Meta may need additional information, including the text of the PLAINTIFFS’ WORKS, in order to provide a more complete response to this Interrogatory. Meta will supplement this response on a rolling basis as it identifies responsive information.

SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 3 (AUGUST 24, 2026):

Subject to and without waiving the foregoing objections, and following a reasonable inquiry, Meta further responds as follows:

Attached as **Exhibit A** hereto is a table identifying, based on Meta’s knowledge, information, and belief based on reasonable investigation, the authors and titles of PLAINTIFFS’ WORKS that Meta believes were downloaded from the Z-Lib, Libgen Fiction, Libgen Scitech and Books3 collections. Based on Meta’s current knowledge, information, and belief based on reasonable investigation, Meta has not identified any PLAINTIFFS’ WORKS that Meta believes were obtained from Common Crawl, an open repository of web crawl data made available by the Common Crawl Foundation. With respect to each of these data sources, Meta identifies, to the best of its knowledge, information and belief after reasonable investigation, how each of these data sources was obtained:

1 **Anna's Archive (which includes Z-Lib):** Meta explained, in its response to Interrogatory
2 No. 9 (served August 8, 2026), the way in which Meta obtained data from the Z-Lib collection of
3 Anna's Archive via torrenting, and that response is incorporated herein by reference.

4 **Libgen Fiction:** Meta explained, in its response to Interrogatory No. 9 (served August 8,
5 2026), the way in which data from Libgen Fiction was obtained via torrenting, and that response is
6 incorporated herein by reference. Additionally, beginning in or about Spring 2023, data from the
7 Fiction branch of Libgen was downloaded, and this download was conducted primarily by Nikolay
8 Bashlykov. The download was conducted using a direct download technique that did not involve
9 use of torrents. Meta has produced a software script (download_libgen_direct.py, META-
10 KADREY-SC-000197), a version of which Meta believes was used in the direct download process
11 for Libgen Fiction.
12

13 **Libgen Scitech:** Meta explained, in its response to Interrogatory No. 9 (served August 8,
14 2026), the way in which data from Libgen Scitech was obtained via torrenting, and that response is
15 incorporated herein by reference. Additionally, beginning in or about Spring 2023, data from the
16 Scitech branch of Libgen was downloaded, and the download was conducted primarily by Nikolay
17 Bashlykov. The download was conducted using a direct download technique that did not involve
18 use of torrents. Meta has produced a software script (download_libgen_direct.py, META-
19 KADREY-SC-000197), a version of which Meta believes was used in the direct download process
20 for Libgen Scitech.
21

22 **Common Crawl:** Meta believes that Common Crawl data was first obtained by Meta prior
23 to 2023. Common Crawl was available for download over the Internet from a number of different
24 websites and online databases. Based on Meta's knowledge and information after a reasonable
25 investigation, Meta has not been able to definitively determine where it originally obtained
26 Common Crawl data, but believes it may have been obtained from
27 <<https://www.tensorflow.org/datasets/catalog/c4>>, <<https://github.com/speedingghzl/CCNet/>>, or
28

1 <[https://github.com/tensorflow/models/tree/archive/research/lm_commonsense#1-download-](https://github.com/tensorflow/models/tree/archive/research/lm_commonsense#1-download-datafiles)
 2 [datafiles](https://github.com/tensorflow/models/tree/archive/research/lm_commonsense#1-download-datafiles)>.

3 **Books3:** Meta believes that Books3 data was first obtained prior to 2023. At the time it
 4 was obtained by Meta, Books3 was part of a larger dataset known as “The Pile” published by
 5 EleutherAI, which was available for download over the Internet from a number of different
 6 websites and online databases. Meta has not been able to definitively determine where it originally
 7 obtained Books3 data, but believes Books3 may have been obtained from a website known as The
 8 Eye (<<https://theeye.eu>>), which at the time hosted a number of publicly-available datasets
 9 including Books3 and The Pile. Meta has produced a spreadsheet (Meta_Kadrey_00000250)
 10 listing names of files in a Books3 dataset obtained by Meta.
 11

12 Meta’s investigation is ongoing and Meta will supplement this Interrogatory as additional
 13 responsive information becomes available.
 14

15
 16 **INTERROGATORY NO. 4:**

17 Which PLAINTIFFS’ WORKS did YOU use, in whole or in part, to train any of YOUR LLM
 18 MODELS?

19 **RESPONSE TO INTERROGATORY NO. 4:**

20 Meta incorporates by reference its objections and definitions above, including to the terms
 21 PLAINTIFFS’ WORKS, YOU, and YOUR LLM MODELS.

22 Meta further objects to this interrogatory to the term “train” as vague and ambiguous to the
 23 extent it is undefined. Meta will interpret the term “train” as the act of pre-training and/or fine-
 24 tuning a large language model.

25 Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable
 26 investigation to identify which PLAINTIFFS’ WORKS, if any, were used to pre-train Meta’s LLM
 27 MODELS. As previously communicated to counsel for PLAINTIFFS, Meta may need additional
 28 information, including the text of the PLAINTIFFS’ WORKS, in order to provide a more complete

1 response to this Interrogatory. Meta will supplement this response on a rolling basis as it identifies
2 responsive information.

3
4 **SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 4 (AUGUST 24, 2026):**

5 Subject to and without waiving the foregoing objections, and following a reasonable
6 inquiry, Meta further responds as follows:

7 Attached as **Exhibit C** is a list of PLAINTIFFS' WORKS that based on Meta's knowledge,
8 information, and belief based on reasonable investigation were used in training one or more of the
9 Llama models. **Exhibit C** is broken into two parts. **Part 1** provides a listing that identifies the
10 work files as produced by PLAINTIFFS (with corresponding title), in which Meta believes the
11 work (or a substantial portion thereof) was used to train at least one of its Llama models. **Part 2**
12 provides a listing that shows, for each work file as produced by PLAINTIFFS, an identification of
13 the training dataset and specific training dataset file, that Meta believes includes the work (or a
14 substantial portion thereof). To the best of Meta's knowledge, information and belief, these
15 training files were used in the training of Meta models and thus provide an identification of specific
16 instances in which the identified PLAINTIFFS' WORKS were used in training.
17

18 Meta's investigation is ongoing and Meta will supplement this Interrogatory as additional
19 responsive information becomes available.
20

21
22 **INTERROGATORY NO. 5:**

23 For each PLAINTIFFS' WORK that YOU used, in whole or in part, to train any of YOUR
24 LLM MODELS, when and how did you do so? IDENTIFY and describe each instance in detail.

25 **RESPONSE TO INTERROGATORY NO. 5:**

26 Meta incorporates by reference its objections and definitions above, including to the terms
27 PLAINTIFFS' WORKS, YOU, and YOUR LLM MODELS.
28

Meta further objects to this interrogatory to the term “train” as vague and ambiguous to the extent it is undefined. Meta will interpret the term “train” as the act of pre-training and/or fine-tuning a large language model. Meta further objects to the phrase “how did you do so?” as it does not specify which actions the interrogatory seeks information about. Meta will interpret the phrase “how did you do so?” as seeking a high-level general description of the way a work like a PLAINTIFFS’ WORK would be used in the pre-training process of Meta’s LLM MODELS. Meta further objects to this interrogatory as vague, ambiguous, overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent the phrase “[i]dentify and describe each instance in detail” seeks unspecified information about the use of PLAINTIFFS’ WORKS in pre-training. Meta will interpret this phrase to seek a high-level general description of the way a work like a PLAINTIFFS’ WORK would be used in the pre-training process of Meta’s LLM MODELS.

Subject to and without waiving the foregoing objections, as stated for Interrogatory No. 4, Meta is undertaking a reasonable investigation to identify which PLAINTIFFS’ WORKS, if any, that were used to pre-train Meta’s LLM MODELS and Meta will provide a high-level general description of the way a work like a PLAINTIFFS’ WORK would be used in the pre-training process of Meta’s LLM MODELS. As previously communicated to counsel for PLAINTIFFS, Meta may need additional information, including the text of the PLAINTIFFS’ WORKS, in order to provide a more complete response to this Interrogatory. Meta will supplement this response on a rolling basis as it identifies responsive information.

SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 5 (AUGUST 24, 2026):

Subject to and without waiving the foregoing objections, and following a reasonable inquiry, Meta further responds as follows:

Meta has provided, in response to Interrogatory 7, *infra*, a description of how textual training data (which would include the PLAINTIFFS’ WORKS identified in response to Interrogatory No. 4) was processed and used in the training process for Meta’s models. As described in response to Interrogatory No. 4, *supra*, **Exhibit C, Part 2**, provides a listing that

1 shows, for each work file as produced by PLAINTIFFS, an identification of the training dataset
 2 and training dataset file, where Meta believes the work (or a substantial portion thereof), is
 3 contained. To the best of Meta's knowledge, information and belief, these training files were used
 4 in the training of Meta models and thus provides an identification of specific instances in which the
 5 identified PLAINTIFFS' WORKS were used in training.

6 Meta's investigation is ongoing and Meta will supplement this Interrogatory as additional
 7 responsive information becomes available.
 8

9
 10 **INTERROGATORY NO. 6:**

11 IDENTIFY each PLAINTIFFS' WORK from which YOU removed or altered any COPYRIGHT
 12 MANAGEMENT INFORMATION.

13 **RESPONSE TO INTERROGATORY NO. 6:**

14 Meta incorporates by reference its objections and definitions above, including to the terms
 15 PLAINTIFFS' WORKS, YOU, and COPYRIGHT MANAGEMENT INFORMATION.

16 Meta further objects to this interrogatory as overbroad, unduly burdensome, and
 17 disproportionate to the needs of the case to the extent it seeks information about PLAINTIFFS'
 18 WORKS that were not used as part of Meta's pre-training of large language models.

19 Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable
 20 investigation to identify for which PLAINTIFFS' WORKS identified in its responses to
 21 Interrogatory Nos. 1 and 2 Meta is able to reasonably determine that COPYRIGHT
 22 MANAGEMENT INFORMATION was removed or altered by Meta. As previously
 23 communicated to counsel for PLAINTIFFS, Meta may need additional information, including the
 24 text of the PLAINTIFFS' WORKS, in order to provide a more complete response to this
 25 Interrogatory. Meta will supplement this response on a rolling basis as it identifies responsive
 26 information.
 27

28 **SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 6 (AUGUST 24, 2026):**

1 Subject to and without waiving the foregoing objections, and following a reasonable
2 inquiry, Meta further responds as follows:

3 As explained in response to Interrogatory No. 7, *infra*, Meta did employ processes in which
4 repetitive text from books was not included in the training data used for training certain Meta Llama
5 models, including data that commonly appear in works such as “ISBN,” “Copyright,” the copyright
6 symbol ©, “All Rights Reserved,” and “DOI.” Based on Meta’s current understanding and
7 investigation, Meta believes this processing step may have been performed in the processing of at
8 least some PLAINTIFFS’ WORKS identified in response to Interrogatory No. 4, *supra*. Meta is
9 undertaking a reasonable search of the produced training data to determine if it can more precisely
10 identify particular PLAINTIFFS’ WORKS to which this processing step was applied. As further
11 explained in response to Interrogatory No. 7, *infra*, Meta did not employ this processing step for
12 scanned and digitized books it used as training data for its Muse models.

14 Meta’s investigation is ongoing and Meta will supplement this Interrogatory as additional
15 responsive information becomes available.

17
18 **INTERROGATORY NO. 7:**

19 Explain what you did with each PLAINTIFFS’ WORK that YOU obtained from a SHADOW
20 LIBRARY or through WEB SCRAPING or that YOU used, in whole or in part, to train any of
21 YOUR LLM MODELS. For each PLAINTIFFS’ WORK, your explanation should state whether,
22 when, and how it was:

23 (a) retained, stored, or preserved;

24 (b) ingested;

25 (c) processed;

26 (d) tokenized;

27 (e) duplicated or reproduced;

28 (f) de-duplicated;

- 1 (g) stripped of COPYRIGHT MANAGEMENT INFORMATION;
- 2 (h) modified or altered in any other way;
- 3 (i) distributed;
- 4 (j) torrented;
- 5 (k) incorporated, in whole or in part, into any of YOUR LLM MODELS or model
- 6 parameters; or
- 7 (l) deleted.

8 **RESPONSE TO INTERROGATORY NO. 7:**

9 Meta incorporates by reference its objections and definitions above, including to the terms
 10 PLAINTIFFS' WORKS, SHADOW LIBRARY, WEB SCRAPING, COPYRIGHT
 11 MANAGEMENT INFORMATION, YOU, and YOUR LLM MODELS.

12 Meta further objects to this interrogatory as compound because it comprises at least twelve
 13 subparts and would potentially exceed the allowable number of interrogatories under the Federal
 14 Rules of Civil Procedure. Meta further objects to the terms "[e]xplain what you did", "processed",
 15 "tokenized", "stripped", "modified or altered in any other way", "incorporated, in whole or in part,
 16 into any of YOUR LLM MODELS or model parameters" as vague and ambiguous. Meta further
 17 objects to this interrogatory to the extent it seeks information sought by Interrogatory Nos. 1-6.
 18 Meta further objects to this interrogatory to the extent it requires Meta to create or re-create
 19 information or data that is not in Meta's possession, custody, or control.

20 Subject to and without waiving the foregoing objections, as stated for Interrogatory Nos. 1-
 21 6, Meta is undertaking a reasonable investigation to identify which PLAINTIFFS' WORKS, if any,
 22 were downloaded and used to pre-train Meta's LLM MODELS. Meta will provide a high-level
 23 general description of the way a work like a PLAINTIFFS' WORK would be utilized by Meta in
 24 conjunction with Meta's LLM MODELS. As previously communicated to counsel for
 25 PLAINTIFFS, Meta may need additional information, including the text of the PLAINTIFFS'
 26 WORKS, in order to provide a more complete response to this Interrogatory. Meta will supplement
 27 this response on a rolling basis as it identifies responsive information.

SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 7 (AUGUST 24, 2026):

Subject to and without waiving the foregoing objections, and following a reasonable inquiry, Meta further responds as follows:

This response is designated as Highly Confidential - Attorneys' Eyes Only under the Protective Order.

As explained in response to Interrogatory No. 3, Meta used publicly-available works that were downloaded from online repositories (in some cases through torrent downloads) and web data, in connection with the research, development, and training of certain AI models. An initial step is to download the files, which are (in the case of books) typically in Portable Document Format (PDF) or Electronic Publication (EPUB) format. Regardless of the format, downloaded files must then be processed into a form suitable for model training. In many instances, the downloaded book files contained a series of scanned pages, which were run through optical character recognition (OCR) to extract textual information from the files. Textual information from the downloaded files was further filtered and processed for quality and de-duplication. In some cases, the processing of the downloaded files resulted in certain text from those files not being included in the data ultimately used for model training. For example, Meta created software scripts to filter out text comprising certain personally identifiable information (PII) (such as email addresses) and repetitive text in book files from certain downloaded data. This repetitive text included (as explained at Meta_Kadrey_00065249), lines that contain fewer than 8% unique words, and repetitive words and phrases that commonly appear in works such as "ISBN," "Copyright," the copyright symbol ©, "All Rights Reserved," and "DOI." *See also* Meta's Response to Interrogatory No. 6, for more specificity. Meta notes that, to the extent other data from the original works was not included in the training data, that fact will be reflected and can be ascertained from the training data produced to Plaintiffs.

1 After the downloading, de-duplication, processing and cleaning processes, the resulting data
2 may then be used as input text for training a large language model (LLM). The LLM training
3 process is a complex one involving numerous steps, including steps of tokenization and
4 embedding. Tokenization generally breaks the input text into smaller units (tokens), which are
5 assigned unique numbers. The embedding process then identifies multi-dimensional vectors
6 (embeddings) for each token in the input sequence, which attempt to capture the meaning of the
7 words in the input text. The training process continues by feeding these word embeddings into the
8 transformer layers of the LLM, to obtain next token probabilities from the neural network, which
9 are compared and used to update the model weights (also commonly referred to as
10 parameters). These training steps are computationally intensive and are repeated billions or even
11 trillions of times (depending on the training dataset size) until the LLM has made predictions about
12 all of the input text in the training dataset. Billions of LLM weights are adjusted throughout this
13 process to improve the model's next-token prediction accuracy. *See* Meta's Response to
14 Interrogatory No. 4, for an identification of the particular PLAINTIFFS' WORKS that Meta
15 believes were used (in whole or part) for model training.

18 Once an LLM has been trained and is available to be used, it can be provided to users in a
19 number of ways. For Llama 1 through Llama 4, Meta made these models available for distribution
20 to the open source community for use by researchers, developers, and other third parties, pursuant
21 to applicable open source license agreements. Although the weights in an open source model are
22 distributed and available to the open source community, the underlying text data used to train the
23 LLM was not provided with these models or their model weights. Beginning in or about April
24 2024, Meta also made functionality from Llama broadly available to the general public through
25 Meta AI and incorporated into certain other Meta products.

27 Beginning in or around mid-2025, Meta began acquiring physical copies of books for use
28 in training later LLMs. Many of the processes described above for models trained on downloaded

works also apply to these models, with some exceptions. The acquired physical copies were first scanned into digital form and then run through OCR to extract their textual data from the scanned images as discussed above, and then de-duplicated, processed and cleaned, but copyright-related information from these purchased books was not filtered out before being used for model training. In April 2026, Meta introduced the initial version of its Muse Spark model, which is available through certain Meta products and services, and later versions were made available to developers through an application programming interface (API). In August 2026, Meta also made its Muse Glimmer model available for distribution to the open source community for use by researchers, developers, and other third parties, pursuant to an open source license agreement.

INTERROGATORY NO. 8:

For each response to every Request for Admission served on YOU by Plaintiff in this case that is not an unqualified admission, IDENTIFY the following information:

- (a) The number of the Request for Admission;
- (b) All facts on which YOU base your response;
- (c) All PERSONS who have knowledge of those facts; and
- (d) All DOCUMENTS that support YOUR response along with the custodian of those DOCUMENTS.

RESPONSE TO INTERROGATORY NO. 8:

Meta incorporates by reference its objections and definitions above, including to the terms YOU and DOCUMENTS.

Meta further objects to this interrogatory as compound because it comprises multiple subparts and would potentially exceed the allowable number of interrogatories under the Federal Rules of Civil Procedure. Meta will respond to this interrogatory to the extent required by Local Rule 33-2 and 36-2.

Meta is willing to meet and confer regarding the appropriate scope, if any, of a response to this interrogatory.

1
2 Dated: August 24, 2026

COOLEY LLP

3
4 By: /s/ Phillip E. Morton

Phillip E. Morton

5
6 Karen L. Dunn (*pro hac vice*)
7 Jessica E. Phillips (*pro hac vice*)
8 Kyle N. Smith (*pro hac vice*)
9 DUNN ISAACSON RHEE LLP
10 401 9th Street NW
11 Washington, DC 20004
12 Telephone: (202) 240-2900
13 Email: kdunn@dirllp.com
14 jphillips@dirllp.com
15 ksmith@dirllp.com

16 Kathleen R. Hartnett
17 Mark R. Weinstein
18 Elizabeth Lee Stameshkin
19 COOLEY LLP
20 3175 Hanover Street
21 Palo Alto, CA 94304
22 Telephone: 650-843-5000
23 Facsimile: 650-849-7400
24 Email: khartnett@cooley.com
25 mweinstein@cooley.com
26 lstameshkin@cooley.com

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

Phillip E. Morton (*pro hac vice pending*)
COOLEY LLP
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004
Telephone: (202) 842-7800
Email: pmorton@cooley.com

Angela L. Dunning

CLEARY GOTTlieb STEEN & HAMILTON
LLP

151 University Avenue

Palo Alto, CA 94301-1632

Telephone: (650) 815-4131

Email: adunning@cgsh.com

Counsel for Defendant Meta Platforms, Inc.

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

Author	Book	Downloaded Dataset(s)
John Carreyrou	BAD BLOOD: Secrets and Lies in a Silicon Valley Startup	Libgen Scitech, Z-Lib, Books3
Philip Shishkin	Restless Valley: Revolution, Murder, and Intrigue in the Heart of Central Asia	Libgen Scitech
Matthew Sacks	Pro Website Development and Operations	Libgen Scitech
Michael S. Kochin	Five Chapters on Rhetoric: Character Action Things Nothing and Art	Libgen Scitech
Michael S. Kochin, Michael Taylor	An Independent Empire: Diplomacy & War in the Making of the United States	Libgen Scitech
Jane Adams	How to Sell What You Write	Libgen Fiction
Jane Adams	Boundary Issues: Using Boundary Intelligence to get the Intimacy you want and the Independence you need in life, love and work	Libgen Scitech

2026-08-24 Cambrome Ex C

Part 1 - List of Works

Produced Book ID	Title
Cambrome_Meta_0000037.txt	Bad Blood searchable (without 2022 afterword)
Cambrome_Meta_0000994.txt	Pro Website Development and Operations
Cambrome_Meta_0001111.txt	Kochin_Revised - review copy - Independent Empire
Cambrome_Meta_0001432.txt	Michael S. Kochin - Five Chapters on Rhetoric_ Character, Action, Things, Nothing, and Art-Penn State University Press (2015) (1)

HIGHLY CONFIDENTIAL - ATTORNEYS EYES ONLY

2026-08-24 Cambronne Ex C

Part 2 - Training Datasets

Produced Book ID	Training Dataset	Training Data File	Title
Cambronne_Meia_0000037.txt	libgen_fiction_20231120	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libfiction_20231120\fiction_en_processed_20231120_chunk_31.jsonl	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	scribbr_en	MEIA_ENTREPRENEUR_DATA_0017schiech_en_union_train_bucket_9.tsv	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	scribbr_en	MEIA_KADREY_DATA_0017schiech_en_union_train_bucket_3.tsv	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	libgen_fiction_20231120	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libscribbr_20231120\scribbr_en_processed_20231120_chunk_08.jsonl	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	b3q	MEIA_KADREY_DATA_001b3q3q_chunk_26.jsonl	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	libgen3_fiction_split	MEIA_ENTREPRENEUR_DATA_001libgen3_fiction_split.tsv	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	libgen_fiction_text_safe	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libfiction_20231210safe\data_chunk_029.jsonl	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000037.txt	scribbr_20231120	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libscribbr_20231120\scribbr_en_processed_20231120_chunk_30.jsonl	Bad Blood searchable (without 2022 afterword)
Cambronne_Meia_0000094.txt	scribbr_20231120	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libscribbr_20231120\scribbr_en_processed_20231120_chunk_02.jsonl	Pro Website Development and Operations
Cambronne_Meia_0000094.txt	scribbr_20231120	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libscribbr_20231120\scribbr_en_processed_20231120_chunk_15.jsonl	Pro Website Development and Operations
Cambronne_Meia_0000094.txt	scribbr_en	MEIA_ENTREPRENEUR_DATA_0017schiech_en_union_train_bucket_5.tsv	Pro Website Development and Operations
Cambronne_Meia_0001111.txt	scribbr_20231120	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libscribbr_20231120\scribbr_en_processed_20231120_chunk_12.jsonl	Kochin Revised - review copy - Independent Empire
Cambronne_Meia_0001432.txt	scribbr_en	MEIA_ENTREPRENEUR_DATA_0017schiech_en_union_train_bucket_3.tsv	Michael S. Kochin - Five Chapters on Rhetoric: Character Action, Things, Nothing, and At-Penn State University Press (2015) (1)
Cambronne_Meia_0001432.txt	scribbr_en	MEIA_KADREY_DATA_0056eag-westfair_lm_v3\data\mnhashdeduped\libscribbr_20231120\scribbr_en_processed_20231120_chunk_15.jsonl	Michael S. Kochin - Five Chapters on Rhetoric: Character Action, Things, Nothing, and At-Penn State University Press (2015) (1)
Cambronne_Meia_0001432.txt	scribbr_en	MEIA_ENTREPRENEUR_DATA_0017schiech_en_union_train_bucket_1.tsv	Michael S. Kochin - Five Chapters on Rhetoric: Character Action, Things, Nothing, and At-Penn State University Press (2015) (1)

HIGHLY CONFIDENTIAL - ATTORNEYS EYES ONLY

CERTIFICATE OF SERVICE

I am a citizen of the United States and a resident of the District of Columbia. I am employed in Washington, District of Columbia. I am over the age of eighteen years, and not a party to the within action. My business address is Cooley LLP, 1299 Pennsylvania Ave NW Ste. 700, Washington, DC 20004. On the date set forth below I served the documents described below in the manner described below:

**DEFENDANT META PLATFORMS, INC.'S FIRST SUPPLEMENTAL RESPONSES AND
OBJECTIONS TO PLAINTIFF'S FIRST SET OF INTERROGATORIES]**

☒ (BY ELECTRONIC MAIL) I am personally and readily familiar with the business practice of Cooley LLP for the preparation and processing of documents in portable document format (PDF) for e-mailing, and I caused said documents to be prepared in PDF and then served by electronic mail to the parties listed below.

on the following part(ies) in this action:

FREEDMAN NORMAND FRIEDLAND LLP	STRIS & MAHER LLP
Kyle Roche (<i>pro hac vice</i>)	Elizabeth Brannen (# 226234)
Devin (Velvel) Freedman (<i>pro hac vice</i>)	John Stokes (# 310847)
Alex Potter (<i>pro hac vice</i>)	Lauren Martin (# 294367)
155 E. 44th Street, Suite 915	17785 Center Court Dr N, Ste 600
New York NY 10017	Cerritos, CA 90703
Email: vel@fnf.law	Email: ebrannen@stris.com
Email: kroche@fnf.law	Email: jstokes@stris.com
Email: apotter@fnf.law	Email: Jason.Morris@ndlf.com
	Christopher M. Rigali (<i>pro hac vice</i>)
	Jacqueline Sahlberg (<i>pro hac vice</i>)
	1717 K St NW Suite 900
	Washington, DC 20006
	Email: crigali@stris.com
	Email: jsahlberg@stris.com

Executed on August 24, 2026, at San Francisco, CA.

/s/ Phillip E. Morton

Phillip E. Morton