

EXHIBIT G

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

COGNELLA, INC.,

Plaintiff,

v.

META PLATFORMS, INC.,

Defendant.

Case No. 3:26-cv-04053-VC

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

Judge: Hon. Vince Chhabria

1 **PROPOUNDING PARTY: PLAINTIFF COGNELLA, INC.**

2 **RESPONDING PARTY: DEFENDANT META PLATFORMS, INC.**

3 **SET NUMBER: ONE**

4 Pursuant to Federal Rules of Civil Procedure Rules 26 and 33, and Local Rule 33 of the
5 United States District Court for the Northern District of California, Defendant Meta Platforms, Inc.
6 (“Meta”) hereby objects and responds as follows to Plaintiff Cognella, Inc.’s (“Plaintiff”) First Set
7 of Interrogatories (“Interrogatories”).

8 **I. Responses to All Interrogatories**

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

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

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

II. Objections to Instructions and Definitions

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

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

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

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

4. Meta further objects to the definitions of “COMMUNICATIONS” and “DOCUMENT(S)” to the extent they seek to expand or alter the scope of the Federal Rules of Civil Procedure, the Local Rules of this Court, or the forthcoming ESI Order. Meta further objects to these definitions to the extent they seek Meta to produce documents that are not reasonably accessible because of undue burden or cost pursuant to Fed. R. Civ. P. 26(b)(2)(B), which are

1 unduly burdensome to preserve and search, or which otherwise, under the proportionality factors,
2 should not be preserved.

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

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

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

27 **8.** Meta objects to the definition of “META,” “YOU,” and “YOUR” as overbroad,
28 unduly burdensome, and disproportionate to the needs of the case. Meta will construe “META,”

1 “YOU,” or “YOUR” to mean Meta Platforms, Inc.

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

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

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

1 MODELS.” Meta will construe “WEB SCRAPING” to mean the automated collection of text data
2 from the public internet for use in training the Meta Language Models (as construed above).

3 **III. General Objections**

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

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

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

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

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

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

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

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

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

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

26 **9.** Meta objects to the Interrogatories to the extent that they seek confidential, trade
27 secret, competitively sensitive, business, financial, or other proprietary information. To the extent
28 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 COGNELLA 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 COGNELLA WORKS, YOU, and SHADOW LIBRARY.

Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable investigation to identify which COGNELLA 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 COGNELLA, Meta may need additional information, including the text of the COGNELLA 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:

Meta will provide a table identifying, based on Meta's current knowledge, information, and belief based on a reasonable investigation, the authors and titles of COGNELLA WORKS, if any, that Meta believes were downloaded from Libgen, Zlib, the Internet Archive Lending Library and/or Books3.

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

INTERROGATORY NO. 2:

Identify each COGNELLA 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 COGNELLA WORK, YOU, and WEB SCRAPING.

Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable investigation to identify which COGNELLA 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 COGNELLA, Meta may need additional information, including the text of the COGNELLA 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):

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

3 Meta will provide a table identifying, based on Meta's knowledge, information, and belief
4 based on reasonable investigation, the COGNELLA WORKS, if any, that Meta believes were
5 obtained from Common Crawl, an open repository of web crawl data made available by the
6 Common Crawl Foundation.

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

11 Meta's investigation is ongoing and Meta will supplement this Interrogatory as additional
12 responsive information becomes available.
13
14
15

16 **INTERROGATORY NO. 3:**

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

20 **RESPONSE TO INTERROGATORY NO. 3:**

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

23 Meta objects to this request to the extent it is duplicative of information sought by
24 Interrogatory Nos. 1 and 2. Meta further objects to this interrogatory as vague, ambiguous,
25 overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent the
26 phrase "[i]dentify and describe each instance in detail, including" seeks unspecified information
27 beyond the recited request for "the source or sources, date or dates of acquisition, and means or
28 method of acquisition" of data obtained from a SHADOW LIBRARY or WEB SCRAPING.

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

11
 12 **SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 3 (AUGUST 24, 2026):**

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

15 Meta will provide a table identifying, based on Meta's knowledge, information, and belief
 16 based on reasonable investigation, the authors and titles of COGNELLA WORKS, if any, that Meta
 17 believes were downloaded from the Z-Lib, Libgen Fiction, Libgen Scitech and/or Books3
 18 collections. Meta will provide a table identifying, based on Meta's knowledge, information, and
 19 belief based on reasonable investigation, the COGNELLA WORKS, if any, that Meta believes were
 20 obtained from Common Crawl, an open repository of web crawl data made available by the
 21 Common Crawl Foundation. With respect to each of these data sources, Meta identifies, to the best
 22 of its knowledge, information and belief after reasonable investigation, how each of these data
 23 sources was obtained:
 24

25 **Anna's Archive (which includes Z-Lib):** Meta explained, in its response to Interrogatory
 26 No. 9 (served August 8, 2026), the way in which Meta obtained data from the Z-Lib collection of
 27 Anna's Archive via torrenting, and that response is incorporated herein by reference.
 28

1 **Libgen Fiction:** Meta explained, in its response to Interrogatory No. 9 (served August 8,
2 2026), the way in which data from Libgen Fiction was obtained via torrenting, and that response is
3 incorporated herein by reference. Additionally, beginning in or about Spring 2023, data from the
4 Fiction branch of Libgen was downloaded, and this download was conducted primarily by Nikolay
5 Bashlykov. The download was conducted using a direct download technique that did not involve
6 use of torrents. Meta has produced a software script (download_libgen_direct.py, META-
7 KADREY-SC-000197), a version of which Meta believes was used in the direct download process
8 for Libgen Fiction.
9

10 **Libgen Scitech:** Meta explained, in its response to Interrogatory No. 9 (served August 8,
11 2026), the way in which data from Libgen Scitech was obtained via torrenting, and that response is
12 incorporated herein by reference. Additionally, beginning in or about Spring 2023, data from the
13 Scitech branch of Libgen was downloaded, and the download was conducted primarily by Nikolay
14 Bashlykov. The download was conducted using a direct download technique that did not involve
15 use of torrents. Meta has produced a software script (download_libgen_direct.py, META-
16 KADREY-SC-000197), a version of which Meta believes was used in the direct download process
17 for Libgen Scitech.
18

19 **Common Crawl:** Meta believes that Common Crawl data was first obtained by Meta prior
20 to 2023. Common Crawl was available for download over the Internet from a number of different
21 websites and online databases. Based on Meta's knowledge and information after a reasonable
22 investigation, Meta has not been able to definitively determine where it originally obtained
23 Common Crawl data, but believes it may have been obtained from
24 <<https://www.tensorflow.org/datasets/catalog/c4>>, <<https://github.com/speedinghzl/CCNet>>, or
25 <[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)
26 [datafiles](https://github.com/tensorflow/models/tree/archive/research/lm_commonsense#1-download-datafiles)>.
27
28

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

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

13 **INTERROGATORY NO. 4:**

14 Which COGNELLA WORKS did YOU use, in whole or in part, to train any of YOUR LLM
15 MODELS?
16

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

18 Meta incorporates by reference its objections and definitions above, including to the terms
19 COGNELLA WORKS, YOU, and YOUR LLM MODELS.

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

23 Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable
24 investigation to identify which COGNELLA WORKS, if any, were used to pre-train Meta’s LLM
25 MODELS. As previously communicated to counsel for COGNELLA, Meta may need additional
26 information, including the text of the COGNELLA WORKS, in order to provide a more complete
27 response to this Interrogatory. Meta will supplement this response on a rolling basis as it identifies
28 responsive information.

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

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

Meta will provide a list of COGNELLA WORKS that based on Meta's knowledge, information, and belief based on reasonable investigation were used in training one or more of the Llama models. To the best of Meta's knowledge, information and belief, these training files were used in the training of Meta models and thus provide an identification of specific instances in which the identified COGNELLA WORKS were used in training.

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

INTERROGATORY NO. 5:

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

RESPONSE TO INTERROGATORY NO. 5:

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

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 COGNELLA 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

1 instance in detail” seeks unspecified information about the use of COGNELLA WORKS in pre-
2 training. Meta will interpret this phrase to seek a high-level general description of the way a work
3 like a COGNELLA WORK would be used in the pre-training process of Meta’s LLM MODELS.

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

12
13 **SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 5 (AUGUST 24, 2026):**

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

16 Meta has provided, in response to Interrogatory 7, *infra*, a description of how textual
17 training data (which would include the COGNELLA WORKS identified in response to
18 Interrogatory No. 4) was processed and used in the training process for Meta’s models. As
19 described in response to Interrogatory No. 4, *supra*, Meta will provide a listing that shows, for each
20 work file as produced by COGNELLA, an identification of the training dataset and training dataset
21 file, where Meta believes the work (or a substantial portion thereof), is contained. To the best of
22 Meta’s knowledge, information and belief, these training files were used in the training of Meta
23 models and thus provides an identification of specific instances in which the identified
24 COGNELLA WORKS were used in training.

25
26 Meta’s investigation is ongoing and Meta will supplement this Interrogatory as additional
27 responsive information becomes available.
28

INTERROGATORY NO. 6:

IDENTIFY each COGNELLA WORK from which YOU removed or altered any COPYRIGHT MANAGEMENT INFORMATION.

RESPONSE TO INTERROGATORY NO. 6:

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

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

Subject to and without waiving the foregoing objections, Meta is undertaking a reasonable investigation to identify for which COGNELLA WORKS identified in its responses to Interrogatory Nos. 1 and 2 Meta is able to reasonably determine that COPYRIGHT MANAGEMENT INFORMATION was removed or altered by Meta. As previously communicated to counsel for COGNELLA, Meta may need additional information, including the text of the COGNELLA 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. 6 (AUGUST 24, 2026):

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

As explained in response to Interrogatory No. 7, *infra*, Meta did employ processes in which repetitive text from books was not included in the training data used for training certain Meta Llama models, including data that commonly appear in works such as "ISBN," "Copyright," the copyright symbol ©, "All Rights Reserved," and "DOI." Based on Meta's current understanding and investigation, Meta believes this processing step may have been performed in the processing of at

1 least some COGNELLA WORKS identified in response to Interrogatory No. 4, *supra*. Meta is
 2 undertaking a reasonable search of the produced training data to determine if it can more precisely
 3 identify particular COGNELLA WORKS to which this processing step was applied. As further
 4 explained in response to Interrogatory No. 7, *infra*, Meta did not employ this processing step for
 5 scanned and digitized books it used as training data for its Muse models.

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

9
 10 **INTERROGATORY NO. 7:**

11 Explain what you did with each COGNELLA WORK that YOU obtained from a SHADOW
 12 LIBRARY or through WEB SCRAPING or that YOU used, in whole or in part, to train any of
 13 YOUR LLM MODELS. For each COGNELLA WORK, your explanation should state whether,
 14 when, and how it was:

- 15 (a) retained, stored, or preserved;
- 16 (b) ingested;
- 17 (c) processed;
- 18 (d) tokenized;
- 19 (e) duplicated or reproduced;
- 20 (f) de-duplicated;
- 21 (g) stripped of COPYRIGHT MANAGEMENT INFORMATION;
- 22 (h) modified or altered in any other way;
- 23 (i) distributed;
- 24 (j) torrented;
- 25 (k) incorporated, in whole or in part, into any of YOUR LLM MODELS or model
- 26 parameters; or
- 27 (l) deleted.

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

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

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

Subject to and without waiving the foregoing objections, as stated for Interrogatory Nos. 1-6, Meta is undertaking a reasonable investigation to identify which COGNELLA WORKS, if any, were downloaded and used to pre-train Meta’s LLM MODELS. Meta will provide a high-level general description of the way a work like a COGNELLA WORK would be utilized by Meta in conjunction with Meta’s LLM MODELS. As previously communicated to counsel for COGNELLA, Meta may need additional information, including the text of the COGNELLA 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. 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,

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

20 After the downloading, de-duplication, processing and cleaning processes, the resulting data
21 may then be used as input text for training a large language model (LLM). The LLM training
22 process is a complex one involving numerous steps, including steps of tokenization and
23 embedding. Tokenization generally breaks the input text into smaller units (tokens), which are
24 assigned unique numbers. The embedding process then identifies multi-dimensional vectors
25 (embeddings) for each token in the input sequence, which attempt to capture the meaning of the
26 words in the input text. The training process continues by feeding these word embeddings into the
27 transformer layers of the LLM, to obtain next token probabilities from the neural network, which
28

1 are compared and used to update the model weights (also commonly referred to as
2 parameters). These training steps are computationally intensive and are repeated billions or even
3 trillions of times (depending on the training dataset size) until the LLM has made predictions about
4 all of the input text in the training dataset. Billions of LLM weights are adjusted throughout this
5 process to improve the model's next-token prediction accuracy. *See* Meta's Response to
6 Interrogatory No. 4, for an identification of the particular COGNELLA WORKS that Meta believes
7 were used (in whole or part) for model training.
8

9 Once an LLM has been trained and is available to be used, it can be provided to users in a
10 number of ways. For Llama 1 through Llama 4, Meta made these models available for distribution
11 to the open source community for use by researchers, developers, and other third parties, pursuant
12 to applicable open source license agreements. Although the weights in an open source model are
13 distributed and available to the open source community, the underlying text data used to train the
14 LLM was not provided with these models or their model weights. Beginning in or about April
15 2024, Meta also made functionality from Llama broadly available to the general public through
16 Meta AI and incorporated into certain other Meta products.
17

18 Beginning in or around mid-2025, Meta began acquiring physical copies of books for use
19 in training later LLMs. Many of the processes described above for models trained on downloaded
20 works also apply to these models, with some exceptions. The acquired physical copies were first
21 scanned into digital form and then run through OCR to extract their textual data from the scanned
22 images as discussed above, and then de-duplicated, processed and cleaned, but copyright-related
23 information from these purchased books was not filtered out before being used for model
24 training. In April 2026, Meta introduced the initial version of its Muse Spark model, which is
25 available through certain Meta products and services, and later versions were made available to
26 developers through an application programming interface (API). In August 2026, Meta also made
27
28

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.

Dated: August 24, 2026

COOLEY LLP

By: /s/ Phillip E. Morton

Phillip E. Morton

Karen L. Dunn (*pro hac vice*)
 Jessica E. Phillips (*pro hac vice*)
 Kyle N. Smith (*pro hac vice*)
 DUNN ISAACSON RHEE LLP
 401 9th Street NW

1 Washington, DC 20004
2 Telephone: (202) 240-2900
3 Email: kdunn@dirllp.com
4 jphillips@dirllp.com
5 ksmith@dirllp.com

6 Kathleen R. Hartnett
7 Mark R. Weinstein
8 Elizabeth Lee Stameshkin
9 COOLEY LLP
10 3175 Hanover Street
11 Palo Alto, CA 94304
12 Telephone: 650-843-5000
13 Facsimile: 650-849-7400
14 Email: khartnett@cooley.com
15 mweinstein@cooley.com
16 lstameshkin@cooley.com

17 Bobby A. Ghajar
18 COOLEY LLP
19 1333 2nd Street, Suite 400
20 Santa Monica, CA 90401
21 Telephone: (310) 883-6400
22 Facsimile: (310) 883-6500
23 Email: bghajar@cooley.com

24 Phillip E. Morton (*pro hac vice pending*)
25 COOLEY LLP
26 1299 Pennsylvania Avenue NW, Suite 700
27 Washington, DC 20004
28 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.

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 Kyle Roche (<i>pro hac vice</i>) Devin (Velvel) Freedman (<i>pro hac vice</i>) Alex Potter (<i>pro hac vice</i>) 155 E. 44th Street, Suite 915 New York NY 10017 Email: vel@fnf.law Email: kroche@fnf.law Email: apotter@fnf.law	STRIS & MAHER LLP Elizabeth Brannen (# 226234) John Stokes (# 310847) Lauren Martin (# 294367) 17785 Center Court Dr N, Ste 600 Cerritos, CA 90703 Email: ebrannen@stris.com Email: jstokes@stris.com 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