

September 11, 2026

Via ECF

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

Re: *Cognella, Inc. v. Meta Platforms, Inc.* [Case No. 3:26-cv-04053-VC]; *Cambronne Inc. et al v. Meta Platforms, Inc.* [Case No. 3:26-cv-03725-VC]; *Chicken Soup for the Soul LLC v. Meta Platforms, Inc.* [Case No. 3:26-cv-02333-VC]

Joint Letter Brief RE: Three C's Motion to Compel Defendant's Interrogatory Responses

Dear Magistrate Judge Hixson:

Plaintiffs Cognella, Inc. (“Cognella”), Cambronne Inc., et al (“Cambronne”), and Chicken Soup for the Soul LLC (“Chicken Soup”) (collectively, the “Three Cs”), and Defendant Meta Platforms, Inc. (“Defendant” or “Meta”), parties to the above-captioned action, submit this letter brief regarding Defendant’s responses and objections to Plaintiffs’ First and Third Set of Interrogatories. Plaintiffs ask the Court to overrule Meta’s objections and to order Meta’s further responses to Interrogatory Nos. 2 and 3-7. The parties have met and conferred and are at an impasse.

I. Plaintiffs’ Position

On the first day of discovery (May 21, 2026), the Plaintiffs served eight interrogatories targeting the most relevant information that could be sought in a case like this: which of the Plaintiffs’ copyrighted works did Meta copy, from where, on how many occasions, and what did it then do with the copies? **Exhibit A** [Plaintiffs’ First Set of Interrogatories]. Despite representing in open court on May 22¹ that Meta would be forthcoming, Meta has evaded these questions. An eleventh-hour supplement skirted key questions. The Three Cs Plaintiffs move to compel.

Plaintiffs first identified these deficiencies in writing on June 30, 2026. **Exhibit B**. Defendant first responded on July 7 and the parties met and conferred on July 8. **Exhibit C**. Having received repeated assurances of a forthcoming supplement, Plaintiffs again conferred telephonically on July 23 and confirming Meta’s agreement to meaningfully supplement its responses within a week’s time in a July 24 email. **Exhibit D**. The parties were asked to meet and confer again regarding the written discovery disputes following discovery conferences before Your Honor on August 5 and August 19. The parties conferred again telephonically for an hour on August 26 as memorialized by email on that same date. **Exhibit E**. Despite finally supplementing its initial interrogatory responses on August 24 (see **Exhibit F** [Chicken Soup], **Exhibit G** [Cognella], and **Exhibit H** [Cambronne]), August 29 (see **Exhibit I** [Cognella]), and again on September 4 (see **Exhibit J** [Chicken Soup], **Exhibit K** [Cognella], and **Exhibit L** [Cambronne]),

¹ Case Management Conference Hearing Transcript, May 22, 2026.

Meta has given answers that evade key issues, often in a sophisticated, but nonetheless evasive manner.

The Three Cs issued targeted discovery: which of Plaintiffs' copyrighted works had Meta obtained; where had Meta obtained them; how Meta used them; whether Meta altered CMI; and what Meta kept even after training. Meta's responses have not provided substantive work-by-work answers. Nor has document production supplied this information.²

For each of the Three Cs Plaintiffs, **Interrogatory 3** asks how and when the copies were made from shadow libraries *or web scraping*, including each source or sources, each date or dates of acquisition, and means or method of acquisition. Meta made no attempt to answer this at all for any of the Three C Plaintiffs' work until the last day of fact discovery. This is at the heart of the case. Although it is well known that copying happened, Meta has avoided saying *when and how* it copied Plaintiffs' works. Because Meta copied Plaintiffs' works in secret, Meta not only possesses this relevant information, *it uniquely possesses it*, and must produce it. Most importantly, *Meta does not answer at all as to web scraping*. Meta should be ordered to do so.

Interrogatory 5 asks what Meta did with the copies for each of the Three Cs: "For each [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." Meta refused to describe any instances at all until the final day of discovery, but even in its last-day supplement, Meta limited its answers to its public-facing models. Meta must answer this interrogatory as to all of its models, and all uses made of Plaintiffs' copyrighted material because its "use" is core information as to infringement and to Meta's attempt at a fair use defense.

Meta deflected **Interrogatories 2**, and **6** in similar ways. **Interrogatory 2** addresses web scraping acquisition, as opposed to downloading. Meta says "go fish," asking *Plaintiffs* to identify where Meta copied its books. Meta must answer for its own conduct. **Interrogatory 6** asks about when Meta performed the specifically illegal action (17 U.S.C. § 1202(b)(1)) of removing copyright information from Plaintiffs' works. Meta insists, even in its most recent supplement, in dodging the question by answering it at a higher level of generality ("repetitive text was not included. . ."). Meta should give a straight answer.

Most importantly, Meta has refused to answer Plaintiffs' **Interrogatory 7**, asking the key question "what did Meta do with Plaintiffs' works after copying them?" This critical question underlies all of the analysis performed in similar cases by the courts in this district. In both *Kadrey v. Meta Platforms* and *Bartz v. Anthropic*, Judge Alsup and Judge Chhabria paid careful attention to the *precise* evidence Plaintiffs are seeking from Meta in these interrogatories. *See Kadrey v. Meta Platforms, Inc.*, 788 F. Supp. 3d 1026, 1047 (N.D. Cal. 2025) (discussing relevance of downloading from shadow libraries), (discussing the relevance of the purpose of the copying), 1048 (discussing the relevance of copies *not* used for training), 1050 (amount of the work copied

² Despite repeated requests, Meta has not produced all file listings of the datasets that contain works acquired from Library Genesis and Anna's Archive that were used to train all its LLMs. Meta's 30(b)(6) witness (Beldock) testified on August 28 that Meta could generate such file lists, yet Meta still has not produced them. Plaintiffs must be able to discover what works Meta obtained from Library Genesis and Anna's Archive that it actually used to train its LLMs, and what works it did not use. Plaintiffs' interrogatories and document requests both require Meta to provide this information and to identify all other datasets it created using works from Library Genesis and Anna's Archive and the uses of each dataset, and produce the file lists for those datasets.

is relevant); *Bartz v. Anthropic PBC*, 787 F. Supp. 3d 1007, 1025 (N.D. Cal. 2025) (discussing the relevance of pirating the copies, including the number of copies pirated), 1026-29 (identifying the relevance of *copies retained after training is complete, and the use to which they are put*) (emphasis supplied), 1031 (“For the pirated library copies, however, Anthropic lacked any entitlement to hold copies of the books at all.”).

Meta continues to avoid answering as to works downloaded that were not used for model training at all (it has never answered this, except argumentatively, in claiming that even copies that were not used for training somehow “further” training) and carefully cabins what answers it does give to certain public-facing models only. To facilitate the required judicial analysis, Meta must answer as for *all* uses it made of Plaintiffs’ works. It must also say what it has retained and what it scraped. The retention of copies after completing training is complete cannot “further” training. An “answer to an interrogatory must be responsive to the question” and “should be complete in itself and should not refer to . . . other documents.” *Logtale, Ltd. v. IKOR, Inc.*, 2013 WL 3967750, at *3 (N.D. Cal. July 31, 2013); A party is “not allowed to rewrite the rog into something narrower than it is,” *BSD Crown, Ltd. v. Amazon.com, Inc.*, 2026 WL 1755357, at *2 (N.D. Cal. June 17, 2026) (Hixson, M.J.).

Meta has repeatedly demanded that Plaintiffs’ particularize their discovery responses work-by-work and rebut every speculation that could be imagined, no matter the difficulty or time involved. For example, Meta has demanded: Responses to a collective 1,072 RFAs (**Exhibit M**), granular week-by-week and month-by-month sales and financial data dating back *decades*, despite Plaintiffs’ assurances that it is seeking statutory, *not* actual damages (*see Exhibit N* at Meta’s Interrogatory No. 19 [Cognella] and No. 15 [Chicken Soup and Cambronne]), and all of the authorized channels through which the Plaintiffs’ asserted works have been made available to the public (whether for download, sale, display, or otherwise) including all websites, subscription services, and retail outlets (e.g. bookstores, by name and location) and dates each of the works was first and last available in such channels. *See Exhibit O* at Meta’s Interrogatory No. 8 [Three Cs]. Meta has also served a collective 858 RFPs to the Plaintiffs, including bizarre requests like “DOCUMENTS sufficient to show the media coverage for each of the ASSERTED WORKs by year in the U.S., from first sale or publication to the present.” [*See e.g. Exhibit P* [Meta’s RFP 118 to Chicken Soup, Meta’s RFP 132 to Cognella, and Meta’s RFP 117 to Cambronne Plaintiffs], and “DOCUMENTS sufficient to identify each PERSON who contributed written material to each ASSERTED WORK and the extent of their contributions.” [*See Exhibit P e.g., Chicken Soup RFP 4*].

To date, Meta has only produced only its tables of works taken, without dates of instances, numbers of copies, or information about what it did specifically with each instance of copying of the Plaintiffs’ works, and no real information about what it scraped.

Interrogatory 13 asked for identification of persons who authorized Meta AI training on copyrighted material. Meta has given an answer limited to certain public-facing models. Plaintiffs are seeking answers as to any AI or LLM’s trained and Meta should answer for all of its models.

Interrogatory 15 asked Meta for information on its deletion of a large amount (9TB) of data on *or after* July of 2023. Meta has provided some information about the deletion in July 2023, but nothing later, and ignored Plaintiffs’ request for information on how this data was acquired. It should be ordered to provide a full answer.

Plaintiffs seek an order compelling Meta to serve complete responses to their Interrogatories numbered **2, 3-7, 13 and 15** and an order overruling Meta’s definitional objections and unilateral interpretations of Plaintiffs’ definitions as described more fully in **Exhibit B**.

Plaintiff's definitions are unambiguous and must stand as defined because they are foundational to the nature of each corresponding discovery request.

II. Meta's Position:

None of 3C's complaints about Meta's interrogatory responses have merit. Meta has already fully addressed the purported concerns in its supplemental responses. 3C's also incorrectly assert that "Meta has not produced file listings of the datasets that contain works acquired from Library Genesis and Anna's Archive that were used to train all of its LLMs." *Supra* n.2. Meta produced these in *Kadrey* and were among the first documents produced to 3C's in this case.

Interrogatory No. 2: The 3C's complaints are disconnected from Meta's responses. Meta did not say "go fish." For Common Crawl materials, for example, Meta undertook extensive efforts to perform a text search of these materials. The result of that search was that, of the more than 3,300 asserted works, only four works were present in Common Crawl. (Cognella: 2 works, Chicken Soup: 2 works, Cambronne: 0 works). Ex. J at pdf p. 48; Ex. K at pdf p. 50; Ex. L at 9:1-4. This was not particularly surprising, as Common Crawl is overwhelmingly a collection of Internet web pages whereas the 3C's asserted works are books. Notably, those four Common Crawl works were **already present** in other books datasets that Meta used to train its models, so the Common Crawl search revealed no additional at-issue works.

SpiderMate provides a similar collection of Web materials, but because of its considerably larger size and composition, it is not text searchable in the same way as Common Crawl – but it is indexed and searchable by website URLs. Meta also offered to conduct URL searches multiple times, but the 3C's never identified any URLs. Meta has explained this issue to 3C's several times, including an explanation in the 9/4 response about the infeasibility of text searching SpiderMate. *E.g.*, Ex. K at 9-10.

Nothing in the record suggests that the 3C's books are available in significant numbers on websites crawled by SpiderMate – let alone that building a new SpiderMate text search system would be likely to identify any non-duplicative books not otherwise found in other at-issue datasets. As noted, SpiderMate (like Common Crawl) is overwhelmingly a collection of Web data (not books), and Meta's searches of Common Crawl data revealed no non-duplicative books. The 3C's decision not to ask Meta to search URLs suggests that they too have no expectation or specific facts suggesting that SpiderMate contains any significant number of their works. It would be vastly disproportional to the needs of the case to require Meta to engineer a new text searching system for SpiderMate data when there is no evidence that non-duplicative copies of 3C's works are likely to exist there and that are not captured in Common Crawl or other at-issue datasets.

Interrogatory No. 3: Meta's responses to Interrogatory No. 3 answer all of the questions posed in the interrogatory: (1) "when and how" Meta obtained the 3C's works from so-called "Shadow Libraries" (e.g., Libgen, Anna's Archive, Books3) and via web crawler data from Common Crawl; (2) the "source or sources"; (3) the "date or dates of acquisition," and (4) the "means or method of acquisition." Meta's response also refers to its August 9, 2026 response to Interrogatory 9, which describes in detail how data was obtained via torrenting from Libgen and Anna's Archive (and the 3C's take no issue with that response). 3C's generalized complaints are again disconnected from the substance of Meta's response. For example, Plaintiffs assert that Meta "does not answer at all as to web scraping," but there is a section labeled "Common Crawl" that explains what Meta knows about the source, method and timing of the acquisition of web crawled data. *E.g.*, Ex. K at 12.

Interrogatory No. 4: The 3C's identify no issues with these responses. Meta assumes it was included in error given that Meta's supplemental responses and associated exhibits identify the particular 3C's works that were used to train each of the Meta models at issue in this case.

Interrogatory No. 5: It is not disputed that Meta also answered this interrogatory, explaining when and how Meta used the 3C's works to train the Llama models. The 3C's complaint relates to Meta's objection to identifying every research or internal model for which Meta may have used the 3C's works, even if that model was just an internal experimental model that was never commercially or publicly available. Meta previously described the use of datasets for internal research purposes and experiments in prior submissions to the Court in the *Kadrey* case. See *Kadrey v. Meta Platforms, Inc.*, Case No. 3:23-cv-03417-VC, Dkt. 543-14 (Pineau Decl.) and Dkt. 543-13 (Kambadur Decl.). AI research is fundamentally iterative and experimental; as researchers create internal experimental models and variants of existing models to study the effects of specific changes or to serve limited purposes, and in many cases, these models are used briefly or for a particular purpose. Requiring Meta to conduct a search for these innumerable research and internal models created over a multi-year period will not change the scope of and is not proportional to the needs of this case. Meta has already identified the 3C's works it downloaded from so-called Shadow Libraries, and identified the Meta LLMs that were trained on those works, thus fixing the group of works that could possibly be at issue in this case from those downloads. Nor could these internal models have any conceivable relevance to damages, as (1) they are not public facing and thus cannot contribute to any purported market dilution and (2) the 3C's have elected to pursue only statutory damages, limiting 3C's to a single award of statutory damages "for all infringements involved in the action, with respect to any one work," 17 U.S.C. 504(c)(1). See also *Desire, LLC v. Manna Textiles, Inc.*, 986 F.3d 1253, 1265, 1271 (9th Cir. 2001) (Congress "limited a plaintiff to a single award no matter how many infringements an individual defendant...committed.").

Interrogatory No. 6: Meta fully answered this interrogatory also, explaining what was done to remove repetitive information like "ISBN," "Copyright," the copyright symbol ©, "All Rights Reserved," and "DOI," from downloaded works before they were used to train its Llama models. Meta does not maintain records specifying, on a work-by-work basis, the works from which this information was removed when creating its text training datasets. Meta nevertheless conducted a broader analysis of the data and explained to the 3C's that, based on that analysis, the substantial majority of their downloaded works used to train Llama models had that type of information removed. *E.g.*, Ex. K. at 20. Meta also explained that this category of information was not removed from works that were physically acquired and scanned for use in training Meta's Muse model. *E.g.*, *id.*

If Plaintiffs demand a work-by-work accounting of which 3C's works had repetitive copyright-related information removed, Meta would need to develop another technical solution to obtain that level of granularity. But the answer to this question is also available to the 3C's, as determining whether particular text was removed from their asserted books can be determined through a comparison of (1) the original text of their books (which the 3C's have), and (2) the text training data that incorporates those books (which Meta produced long ago in discovery). Such a comparison would show the extent (if any) of removal of text from their original works.

Interrogatory No. 7: Again, Meta answered this interrogatory with four pages of explanation describing the life cycle of a work from its download to its use in training, describing to the best of Meta's ability, how books like the 3C's works were processed and used to train Meta's LLMs. *E.g.*, Ex. K at 22-25. Meta also made clear in its interrogatory response that all of the 3C's works that were obtained by Meta were used for the purposes of research, development, and training its

AI models and that each of the steps described in the interrogatory was applicable to those works (with an exception regarding the removal of repetitive text in works that were purchased and scanned). *Id.*

The 3C's suggestion that they are entitled to an individualized explanation for every work or every copy of every work is without merit. Judge Chhabria addressed similar arguments in *Kadrey* and rejected them, finding that downloading more works than were used to train is still part of Meta's transformative fair use. *Kadrey*, 788 F. Supp. 3d at 1048. Judge Chhabria also made clear that "fair use doesn't require that the secondary use make the lowest number of copies possible." *Id.* (*citation omitted*). The detailed accounting that the 3C's appear to request, therefore, is not relevant or proportional to the needs of this case.

Interrogatory No. 13: Meta responded to this Interrogatory on August 31 and heard nothing from 3Cs before receiving this letter brief less than 36 hours before the discovery motion deadline. Plaintiffs' request should not be entertained given 3C's failure to meet and confer with Meta per the local rules. N.D. Cal. L.R. 37-3(a). Regarding the substance of the 3C's complaints, for the same reasons as discussed for Interrogatory No. 5, it is not proportional the needs of the case to require Meta to investigate innumerable research and internal models to track down every person who may have approved the acquisition or use of books for training.

Interrogatory No. 15: Again, Plaintiffs did not meet and confer with Meta regarding this response and the motion should not be entertained on that basis. N.D. Cal. L.R. 37-3(a). Regarding the substance, the 3C's request for information "after" July 2023 makes no sense because the request concerns a single event that is described in extensive detail. Regarding Plaintiffs' request about "how this data was acquired," the response states that the data was associated with former Meta employee Guillaume Lample. If necessary, Meta can supplement to make clear that Meta believes that the data was acquired by Guillaume Lample via torrent earlier in 2023 (an acquisition Meta has fully described in its response to the 3C's Interrogatory No. 9).

Exhibit B: 3C's attempt to evade the page limits by incorporating by reference a 7 page letter into this brief. To the extent Plaintiffs have specific concerns with Meta's objections, they should be addressed in the text of this brief. The only objection that seems pertinent to the issues here is the definition of "LARGE LANGUAGE MODELS" or "LLM MODELS", which Meta has addressed above with respect to Interrogatory No. 5 and 13.

Dated: September 11, 2026

By: /s/ Phillip Morton

COOLEY LLP

Bobby A. Ghajar (198719)
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400
Facsimile: (310) 883-6500
Email: bghajar@cooley.com
cghazarian@cooley.com

Mark R. Weinstein (193043)
Elizabeth Lee Stameshkin (260865)
3175 Hanover Street Palo Alto, CA 94304
Telephone: 650-843-5000
Facsimile: 650-849-7400
Email: mweinstein@cooley.com
lstameshkin@cooley.com

Kathleen R. Hartnett (314267)
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111-4004
Telephone: (415) 693-2071
Facsimile: (415) 693-2222 Email:
khartnett@cooley.com

Phillip E. Morton (*pro hac vice*)
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004-2400
Telephone: (202) 728-7055
pmorton@cooley.com

DUNN ISAACSON RHEE LLP

Karen L. Dunn (*pro hac vice*)
Jessica E. Phillips (*pro hac vice*)
Kyle N. Smith (*pro hac vice*)
401 9th Street NW Washington, DC 20004
Telephone: (202) 240-2900
Email: kdunn@dirllp.com
jphillips@dirllp.com
ksmith@dirllp.com

Respectfully submitted,

By: /s/ Lauren Martin

STRIS & MAHER LLP

Elizabeth Brannen (SBN 226234)
ebrannen@stris.com
John Stokes (SBN 310847)
jstokes@stris.com
Lauren Martin (SBN 294367)
lmartin@stris.com
17785 Center Court Dr N, Ste 600
Cerritos, CA 90703
T: (213) 995-6800
F: (213) 261-0299

Christopher M. Rigali (*pro hac vice*)
crigali@stris.com
Jacqueline Sahlberg (*pro hac vice*)
jsahlberg@stris.com
1717 K St NW Ste 900
Washington, DC 20006
T: (202) 800-5749

**FREEDMAN NORMAND
FRIEDLAND LLP**

Kyle W. Roche (*pro hac vice*)
kroche@fnf.law
Devin (Velvel) Freedman (*pro hac vice*)
forthcoming)
vel@fnf.law
155 E. 44th Street, Ste 915
New York, NY 10017
T: (646) 494-2900

Attorneys for Plaintiffs

**CLEARY GOTTlieb STEEN &
HAMILTON LLP**

Angela L. Dunning (212047)

151 University Avenue

Palo Alto, CA 94301-1632

Telephone: (650) 815-4131

Email: adunning@cgsh.com

Attorneys for Defendant Meta Platforms, Inc.

FILER'S ATTESTATION

I, Christopher Regan, am the CM/ECF user whose ID and password are being used to file this Joint Letter Brief. Pursuant to Civil Local Rule No. 5-1(i)(3), I attest that concurrence in the filing of this document has been obtained from each of the other signatories.

Dated: September 11, 2026

By /s/ Christopher Regan
Christopher Regan