1

2

3

4   UNITED STATES DISTRICT COURT

5   NORTHERN DISTRICT OF CALIFORNIA

6

7   RICHARD KADREY, et al.,   Case No. 23-cv-03417-VC   (TSH)

8   Plaintiffs,

9   v.   **PUBLIC VERSION OF DISCOVERY
                                                      ORDER AT ECF NO. 351**
10   META PLATFORMS, INC.,   Re: Dkt. Nos. 308, 309, 321, 334, 335, 336

11   Defendant.

12

13   The Court addresses ECF Nos. 308, 309, 321, 334, 335 and 336 as follows.

14   **A.   ECF No. 308**

15   **1.   RFP 118**

16   Plaintiffs' RFP 118 requested "[a]ll Documents and Communications, including source

17   code, relating to any efforts, attempts, or measures implemented by Meta to prevent Llama Models

18   from emitting or outputting copyrighted material." In the joint discovery letter brief, Plaintiffs

19   move to compel the data mentioned in Sections 3 and 4.2 of Meta's Llama 2 paper and the data

20   mentioned in Sections 4.2 and 5.4.3 of Meta's Llama 3 paper. At the hearing, Plaintiffs made

21   clear they want certain particular data referred to in those sections, not all of it. Accordingly, as

22   discussed at the hearing, the Court **ORDERS** the parties to file a supplemental joint discovery

23   letter brief concerning this RFP by December 23, 2024.[1]

24   **2.   RFP 119**

25   Plaintiffs' RFP 119 requested "[a]ll Documents and Communications, including source

26   code, relating to the processing of copyrighted material used in training Llama Models, including

27

28   _____
     [1] At the hearing Plaintiffs explained that the particular dispute concerning the organization of
     Meta's source code that they raised in ECF No. 308 is no longer live.

United States District Court
Northern District of California

1   storage and deletion of copyrighted material."

2   Plaintiffs move to compel Meta to (1) produce or identify all copies it made of copyrighted

3   works, including but not limited to Plaintiffs' works, (2) search the custodial files of its 15

4   custodians, plus relevant non-custodial databases (which they contend include work email,

5   Workplace Chat and WhatsApp) for documents involving Llama and (a) the removal of copyright

6   management information from literary works or (b) torrenting of data that includes literary works.

7   There are a number of problems with this motion to compel. The first portion of it (all

8   copies of copyrighted works) seeks things not requested by this RFP, which concerns "the

9   processing of copyrighted material used in training Llama Models." The Court also does not see

10   how all copies of copyrighted works is proportional to the needs of this case, which is about the

11   use of copyrighted materials to train the Llama models, not all copyright infringement committed

12   by Meta. During the hearing Plaintiffs offered to narrow their request to datasets that include

13   Plaintiffs' copyrighted works (rather than everyone's copyrighted works), but that limitation did

14   not change the fact that the RFP did not ask for that, or that datasets not used to train the Llama

15   models are not relevant or proportional to the case.

16   With respect to the second part of this motion to compel, the Court does not agree that

17   work email, Workplace Chat and WhatsApp should be treated as non-custodial sources. That

18   would effectively blow up the custodial limitations in the ESI Order (ECF No. 101). Plaintiffs are

19   candid that this is exactly what they are trying to do. They want email searches done for everyone

20   at Meta who works in AI, which Plaintiffs say is a thousand people and Meta says is two

21   thousand. The Court would never have granted that relief no matter when Plaintiffs asked for it,

22   but asking for the number of custodians to be increased from 15 to a thousand (or two thousand)

23   nine days before the close of fact discovery leaves one to wonder what is going on here.

24   However, for the 15 custodians the Court agrees that custodial files regarding Llama and

25   removing copyright management information from literary works are responsive to this RFP and

26   relevant and proportional to the needs of the case. The removal of CMI is relevant to willfulness,

27   for example. The Court does not see how torrenting is responsive to this RFP, which is about the

28   processing of data, not its acquisition. As noted below, the Court does think torrenting is relevant

United States District Court
Northern District of California

2

1   to the case, but it is not responsive to the RFP Plaintiffs moved to compel on.

2   Accordingly, the Court **ORDERS** Meta to search the custodial files for its 15 custodians

3   and produce documents and communications regarding Llama and stripping or removal of CMI

4   from literary works. The Court limits this order to custodial documents because the only alleged

5   non-custodial sources Plaintiffs referred to in the joint discovery letter brief are ones the Court

6   declines to treat as non-custodial. The Court otherwise **DENIES** Plaintiffs' motion to compel.

7   **B.   ECF No. 309**

8   Plaintiffs argue that Meta has improperly redacted non-privileged business-related

9   communications. Plaintiffs argue (correctly) that advice from in-house counsel concerning

10   business matters is not privileged. Plaintiffs also seem to argue that the "primary purpose" test has

11   to be measured against a document as a whole, and that the attorney-client privilege cannot apply

12   to part of a document. But that's not correct. "That the document as a whole addresses

13   predominantly business matters does not negate the privilege as to the portion containing requests

14   for legal advice." *United States v. Chevron Corp.*, 1996 WL 444597, *2 (N.D. Cal. May 30,

15   1996). "Thus, despite the overall nature of the document, the client may assert the attorney-client

16   privilege over isolated sentences or paragraphs within a document." *Id.*; *see also In re Meta Pixel*

17   *Healthcare Litigation*, 2024 WL 3381029, *6 (N.D. Cal. July 10, 2024) (approving redactions for

18   attorney-client privilege).

19   The Court ordered Meta to submit the documents in question for *in camera* review.

20   Having performed an *in camera* review, the Court **SUSTAINS** Meta's privilege claims as to

21   Meta-Kadrey_00146534, Meta-Kadrey_00146557 and Meta-Kadrey_00146583. With respect to

22   Meta-Kadrey_00152812, the Court **SUSTAINS** Meta's privilege claim as to the redaction on

23   Meta-Kadrey_00152826 but **OVERRULES** Meta's privilege claim as to the redaction on Meta-

24   Kadrey_00152834, which describes an action the legal department took and does not contain legal

25   advice. For the same reason, in Meta-Kadrey_00152994, the Court **OVERRULES** Meta's

26   privilege claim as to the reaction on Meta-Kadrey_00153003. The Court **SUSTAINS** Meta's

27   privilege claims as to Meta-Kadrey_00153393, Meta-Kadrey_00153799, Meta-Kadrey_00154472,

28   Meta-Kadrey_00154479 and Meta-Kadrey_00154729. The Court **OVERRULES** Meta's

United States District Court
Northern District of California

3

1   privilege claim as to the redaction in Meta-Kadrey_00155464. Nothing in the document or Meta's

2   privilege log indicates that the referenced approval was not a business decision. The language

3   about legal issues appears to be the author's personal opinion, and Meta does not say in its

4   privilege log that the author is an attorney. Meta's privilege log says that this redaction reflects

5   legal advice, but the Court disagrees. The Court **SUSTAINS** Meta's privilege claim as to Meta-

6   Kadrey_00155715. For Meta-Kadrey_00156178, the Court **SUSTAINS** Meta's privilege claim

7   for the first bullet point under "Initial Convo" on Meta-Kadrey_00156185 but **OVERRULES**

8   Meta's claim of privilege for the redactions beneath that on that page because they are a summary

9   of what TripAdvisor said. The Court **SUSTAINS** Meta's claim of privilege as to page Meta-

10   Kadrey_00156189.

11   **C.   ECF No. 321**

12   **1.   Search Terms**

13   Plaintiffs propose to add five search strings. However, Meta is correct that Plaintiffs'

14   proposal does not comply with the ESI Order (ECF No. 101). A consequence of that is that the

15   Court does not know if these search strings are any good. The ESI Order provides: "If the search

16   terms proposed by the Requesting Party have an unreasonably high or overbroad yield, the

17   Producing Party may review a randomly generated 95/5 confidence level/margin of error sample

18   set of documents to determine the overbreadth of the proposed search terms. Where appropriate,

19   the Producing Party may develop alternative search terms that are more narrowly tailored to

20   capture the relevant, responsive, non-privileged documents from the additional Null Set Sample,

21   and provide a hit report on those terms." The idea is that the search terms are supposed to be

22   informed by whether they have an unreasonably high or overbroad yield, but that analysis hasn't

23   happened yet. Plaintiffs are asking the Court to order Meta to use certain search strings in an

24   informational vacuum.

25   Plaintiffs complain that Meta disclosed its search terms six weeks after Plaintiffs requested

26   them, and more generally that Meta was slow and unresponsive during meet and confer. Meta

27   disagrees. If Plaintiffs thought Meta was acting too slowly, the better course of action would have

28   been to file a discovery letter brief asking the Court to order Meta to disclose its search terms

1   sooner, and more generally to speed up the process under the ESI Order. Instead of doing that,

2   Plaintiffs seek to bypass any vetting of their proposed search terms. The Court is not going to do

3   that. Plaintiffs' motion to compel the use of their search terms is **DENIED**.

4   **2.   Other Issues**

5   As discussed above, Meta's email and Workplace chat are custodial sources, and the Court

6   rejects Plaintiffs' argument to the contrary. The Court will not effectively abolish custodial

7   limitations, as Plaintiffs request.

8   There is nothing wrong with Meta collecting WhatsApp messages from those custodians

9   who said they may have relevant messages.

10   The Court once again declines to expand the time frame for document production by a year

11   and a half in response to a joint discovery letter brief filed just a few days before the close of fact

12   discovery.

13   Plaintiffs' motion to compel is **DENIED** as to these issues.

14   **D.   ECF No. 334**

15   **1.   Clawed Back Documents**

16   Meta clawed back pages 93506 and 93507 from Meta_Kadrey_00093499 during a

17   deposition, then clawed back the same pages from four duplicate versions of the document

18   (Meta_Kadrey_00079969, Meta_Kadrey_00093389, Meta_Kadrey_00093430 and

19   Meta_Kadrey_00093446). Following *in camera* review, the Court **SUSTAINS** the claw back.

20   Those pages contain legal advice and are privileged. And there is no reason to think that the

21   production of these documents without those pages redacted was anything other than inadvertent.

22   **2.   Meta's Decision to Stop Licensing Efforts for Llama Text Data**

23   Meta's Sy Choudhury testified that the decision to pause licensing efforts for Llama text

24   data was made in a meeting he had in April 2023 with his boss Marc Shedroff and Meta's attorney

25   Natascha Parks. Subsequent declarations indicate it was two meetings and the attorney was

26   Morvarid Metanat. In any event, Choudhury acknowledged that "[i]t was a multifaceted decision

27   of which part of this included technical concerns, business concerns, and legal concerns both for --

28   you know, and so it was a multifaceted decision." However, several times during his depositions

1   he claimed that the entire meeting was attorney-client privileged. It is true that in other portions of

2   his depositions he did go into some of the business reasons for the pause. But the Court is

3   concerned that the witness's testimony about the business reasons for the pause may be

4   incomplete because the witness repeatedly claimed that the whole meeting was privileged.

5   Meta is entitled to invoke the attorney-client privilege to shield the legal advice requested

6   or received at that meeting (or meetings), but that's it. Where, as here, a business decision was

7   made for a combination of legal and business reasons, the business reasons are not privileged.

8   Putting a lawyer in a meeting does not make everything privileged. The Court **GRANTS**

9   Plaintiffs' motion to compel and **ORDERS** Meta to make Choudhury available for an additional

10   two hours of deposition, one hour as a 30(b)(1) witness and one as a 30(b)(6) witness, regarding

11   Meta's April 2023 decision to pause or cease Meta's licensing efforts for text data for use with

12   Llama.

E.   ECF No. 335

14   Under the existing pleadings, torrenting is relevant because that's how Meta acquired

15   LibGen. Plaintiffs have not shown that Meta's 30(b)(6) witness on this topic (Michael Clark) was

16   unprepared to testify about torrenting. Plaintiffs include a subheading in the joint discovery letter

17   brief that says: "No Meta Witnesses Were Adequately Prepared to Testify About Torrenting."

18   However, in reviewing the argument they make, their actual complaint is that no witnesses were

19   prepared to testify about seeding, not torrenting.

20   The Court understands from the cited testimony that a requirement of torrenting is that the

21   participant also agree to seeding (i.e., you can't just take data, you have to supply some in return).

22   Plaintiffs have shown that Meta's 30(b)(6) witnesses were unprepared to discuss Meta's seeding in

23   any level of detail. However, the Court does not see that seeding falls within any of Plaintiffs'

24   30(b)(6) deposition topics. The Court also does not see how details about seeding are relevant to

25   the claims in the existing pleadings. Accordingly, Plaintiffs' motion to compel is **DENIED**.

F.   ECF No. 336

27   Plaintiffs contend that Meta is abusing the attorney-client privilege with respect to

28   redactions made to documents concerning its mitigation efforts. Plaintiffs challenge the redactions

1 Meta has made to a number of documents listed in Exhibit C to the joint discovery letter brief.

2 Meta has submitted most of those documents for *in camera* review (it appears that

3 Meta_Kadrey_00051427 was omitted). The Court has reviewed about half of these documents *in*

4 *camera*, which was a time consuming task. In the documents the Court reviewed, the Court saw

5 no problems with Meta's privilege redactions. The Court concludes that further *in camera* review

6 is unwarranted, as there is no indication that Meta is abusing the privilege. The Court also does

7 not see how Meta's restrained invocation of the privilege gives rise to a plausible "sword and

8 shield" problem. Plaintiffs' motion is **DENIED**.

9   **IT IS SO ORDERED.**

10

11 Dated: December 20, 2024

12

13 THOMAS S. HIXSON
United States Magistrate Judge

United States District Court
Northern District of California

7