

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
EUREKA DIVISION

PAUL TREMBLAY, et al.,
Plaintiffs,
v.
OPENAI, INC., et al.,
Defendants.

Case No. 23-cv-03223-AMO (RMI)

ORDER ON DISCOVERY MOTIONS

Re: Dkt. Nos. 237, 242, 244, 246

Before the court are several discovery disputes. Pursuant to Federal Rule of Civil Procedure 78(b) and Civil Local Rule 7-1(b), the court finds the matters suitable for disposition without oral argument.

Letter Briefs Dkts. 237 & 244

The court finds Plaintiffs' requests regarding the noticed depositions to be without merit. The depositions shall take place prior to the end of January, but the court will deny Plaintiffs' request to enter an order requiring them to occur prior to January 17, 2025. As to the deposition protocols, the court does not see a basis for Plaintiffs' requests for the court to enter an order similar to the one entered in the Southern District of New York. The admonition given this past summer was for the parties to streamline and coordinate discovery with the other cases, not mirror them. Considering the fast-approaching discovery deadline and Plaintiffs' failure to show why more than 15 depositions would be necessary, the court will instead adopt the Defendants' proposals as modified. Accordingly, it is **ORDERED**:

(1) that the parties are directed to informally coordinate with the SDNY plaintiffs on scheduling depositions of OpenAI witnesses so that SDNY plaintiffs may attend the depositions if possible; (2) each side is permitted up to 105 hours of fact

depositions (the equivalent of 15 depositions) including party and non-party witnesses, but excluding 30(b)(6) testimony - any additional time required for translation into English will not count against the total cap set forth above; (3) the parties may take up to 20 hours of Rule 30(b)(6) testimony; and (4) “apex” depositions are limited to 3.5 hours.

Should the parties decide to stipulate to more specifics, they are free to do so. Otherwise, the Federal Rules and the court’s Local Rules govern.

Letter Brief Dkt. 242

Plaintiffs’ requests to compel contained within this letter brief are all DENIED. As to RFP No. 31 (Strategic Growth and Business Plans), OpenAI has already agreed to produce “audited financial statements and numerous internal corporate strategy documents” and “perform[] a reasonable search for non-privileged documents discussing or describing (1) how ChatGPT and relevant models can be or are used for OpenAI commercial uses and applications and (2) OpenAI’s decision to integrate ChatGPT or relevant models in OpenAI’s commercial products.” (Dkt. 242 at 4). Plaintiffs fail to show why such production is not sufficient for the needs of this case, nor have they shown why the broad request of all “documents concerning any SWOT analysis, business plan, go-to-market strategy, competitive analysis, executive summary, growth plan, product integration strategy, or revenue operations plan, concerning any of the OpenAI Language Models or ChatGPT” is relevant and proportional beyond what has been and is already being produced.

As to RFP No. 41 (Projected Income), Defendants are correct that damages in copyright actions are limited and that Plaintiffs have failed to provide an adequate basis for documents related to “projected income.” This is especially true considering OpenAI’s production of documents related to actual income and its response to RFP No. 31. Plaintiffs’ claim for the documents under “fair-use factor one” fails for the same reason. As stated above, OpenAI has already agreed to produce “non-privileged documents discussing or describing (1) how ChatGPT and relevant models can be or are used for OpenAI commercial uses and applications and (2) OpenAI’s decision to integrate ChatGPT or relevant models in OpenAI’s commercial products.” This production is sufficient and proportional.

As to RFP No. 45 (Use of Torrents), Plaintiffs concede that OpenAI did not and is not excluding documents containing the term “torrent” while searching for “non-privileged documents

1 discussing the acquisition of text training data used to train the relevant models” and that it has
2 produced some documents responsive to the request. *Id.* at 3. However, Plaintiffs contend that it
3 “appears that other types of these documents have not been produced, such that a narrow, targeted
4 request for this category of documents is warranted and proportional to the needs of the case.” *Id.*
5 Plaintiffs then allude to “gaps” in Defendants’ production as a basis for a targeted ESI search. The
6 court finds the bases for this request to be speculative and unsupported. Plaintiffs have failed to
7 show why the current production is insufficient.

8 As to RFP No. 68 (Documents Regarding Processing Copyrighted Material), based on
9 Defendants’ offer to produce responsive source code and “to perform a reasonable search for non-
10 privileged documents (1) relevant to how OpenAI curates and uses different types of data to train
11 OpenAI’s relevant models and (2) relevant to the methods it implements to avoid copyright
12 infringement for the relevant models” and confirmation “that it is performing a reasonable search
13 for non-privileged documents discussing the deletion of the specific datasets referenced by
14 Plaintiffs,” the court finds that this request is premature. If Plaintiffs are unsatisfied with the
15 production, they may seek the court’s assistance after *meaningful* collaboration with opposing
16 counsel pursuant to the court’s letter brief procedures.

17 ***Letter Brief Dkt. 246***

18 This Letter Brief is related to the oral argument held on December 17, 2024. At that
19 hearing, the court instructed that discovery into the recent and in-development large language
20 models appeared reasonable; but in balancing proportionality at to the needs of the case, the court
21 instructed Plaintiffs to narrow their requests. The parties were then to engage in a back-and-forth
22 and see if they could resolve the dispute without court intervention. These efforts proved
23 unsuccessful. Based on a review of the original requests, the hearing transcript, and the current
24 briefing, the court finds that Defendants’ offer of compromise is appropriate, as it allows for
25 sufficient relevant discovery while properly balancing the burden of production. Plaintiffs have
26 failed to show why the compromise is insufficient. Accordingly, OpenAI has agreed and will be
27 **ORDERED** as follows:

28 (1) its search for complaints about copyright issues, its documents concerning efforts to

1 prevent regurgitation of training materials, and its productions concerning agreements and
2 negotiations with third parties will *not exclude* documents because they concern a GPT-
3 class model in development;

4 (2) it will produce for inspection the text pre-training data for in-development, text-based
5 GPT-class of models where the pre-training phase has been completed and the model is
6 intended for production, including the next GPT-class model still being developed, which
7 has been referred to as “Orion”; and

8 (3) should Plaintiffs have questions about the origin of a particular dataset, or whether that
9 set was legally licensed, they may address them to OpenAI.

10 All other requests with respect to this Letter Brief are **DENIED**.

11 **IT IS SO ORDERED.**

12 Dated: January 13, 2025

13
14 

15 ROBERT M. ILLMAN
16 United States Magistrate Judge
17
18
19
20
21
22
23
24
25
26
27
28