

1
2
3
4 UNITED STATES DISTRICT COURT
5 NORTHERN DISTRICT OF CALIFORNIA
6 EUREKA DIVISION
7

8 PAUL TREMBLAY, *et al.*,

9 Plaintiffs,

10 v.

11 OPENAI, INC., *et al.*,

12 Defendants.

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

**ORDER RE: PLAINTIFFS' MOTION
TO COMPEL PRODUCTION FROM
NON-PARTY REUTERS NEWS AND
MEDIA, INC.**

Re: Dkt. No. 314

13
14 Now pending before the court is Plaintiffs' Motion (dkt. 314) to Compel Production of
15 Documents from Non-Party Reuters News and Media, Inc. (hereafter, "Reuters"); Reuters has
16 responded (dkt. 348); and, Plaintiffs have filed a reply (dkt. 367). Having reviewed the Parties'
17 submissions, pursuant to Federal Rule of Civil Procedure 78(b) and Civil Local Rule 7-1(b), the
18 court finds the matter suitable for disposition without oral argument. For the reasons stated below,
19 Plaintiffs' motion is denied.

20 Plaintiffs begin by contending that in order to prosecute this case and to respond to
21 OpenAI's fair use defense, they require information from Reuters that they cannot get from
22 OpenAI. *See* Pls.' Mot. (dkt. 314) at 2. Plaintiffs believe that "Reuters possesses information
23 relevant to the existence of a market for copyrighted content as training data for LLMs [because]
24 [o]n October 24, 2024, it was reported that Reuters had entered into a multi-year licensing
25 agreement with Meta Platforms Inc. for its news content to be used in Meta's AI ChatBot [which,
26 according to a blog referenced by Plaintiffs, is] one of only a handful of AI data licensing
27 agreements in the LLM training data market." *Id.* In essence, Plaintiffs seek to obtain the
28 agreement between Reuters and Meta Platforms, Inc. (hereafter, "Meta") because of the notion

1 that those documents are relevant to OpenAI’s anticipated fair use defense, “namely the effect of
 2 [OpenAI’s] theft of copyrighted material on the market for copyrighted content as training data for
 3 LLMs, and to assist in valuing the copyrighted works for use as training data, including
 4 developing a damages methodology.” *Id.* More specifically, Plaintiffs’ subpoena seeks three
 5 categories of documents from Reuters: (1) executed AI training data licensing agreements; (2) AI
 6 training data licensing agreement negotiations and valuations; and, (3) specific content for AI
 7 training. *Id.* at 3. Plaintiffs argue that Reuters has resisted compliance with their subpoena
 8 demands through meritless boilerplate objections, many of which Plaintiffs characterize as
 9 “frivolous.” *Id.* at 4.

10 As to relevance, Plaintiffs submit that prying into the confidential business affairs of
 11 Reuters is justified here for two principal reasons: (1) their reported relevance to Plaintiffs’ claims
 12 of direct copyright infringement by OpenAI; and, (2) their importance and necessity for countering
 13 OpenAI’s fair use defense. *Id.* Regarding the first reason, Plaintiffs state that “[t]he documents
 14 sought from Reuters via the subpoena will help establish the existence of a market for copyrighted
 15 work as training data for LLMs because they will provide evidence of benchmark licensing
 16 agreements and market demand¹ . . . [and that the] information is also relevant to calculating
 17 actual damages and recovery of lost profits.” *Id.* at 4-5. In short, Plaintiffs essentially ask the
 18 undersigned to order Reuters to divulge the details of its confidential business arrangements with
 19 Meta simply so that Plaintiffs can show that they “would have [otherwise] been able to license
 20 their works at a competitive price.” *Id.* at 5. Plaintiffs contend that Reuters’ “internal valuation
 21 memos can provide evidence of potential lost revenue due to infringement and would likely
 22 contain projections and analyses of potential licensing income” which might “support their claims
 23 for lost profits.” *Id.* Plaintiffs add that the negotiations surrounding the Reuters/Meta licensing
 24 deal might “provide evidence of the license’s value, helping to quantify [Plaintiffs’] losses [in this
 25 case].” *Id.*

26 As to Plaintiffs’ efforts to counter OpenAI’s anticipated fair-use defense, Plaintiffs contend

27
 28 ¹ It should be noted that Plaintiffs seek this information for news content, which is different from the content involved in this case.

1 that this discovery would relate to “the fourth factor in establishing such defense (the effect of the
2 use upon the potential market for or value of the copyrighted work)[,] [which] requires
3 consideration of whether the unlicensed use of the work undermines the market or licensing
4 opportunities for it.” *Id.* at 5. In other words, Plaintiffs’ argue that “[i]f a licensing market exists
5 and is negatively impacted by OpenAI’s theft of [Plaintiffs’] data, [which] weighs against a
6 finding of fair use,” that the opening up of the files documenting the business relationship between
7 Reuters and Meta “will aid Plaintiffs in this analysis.” *Id.* at 5-6. It is, of course, unclear why
8 Plaintiffs believe that the particulars of their non-news content case will be materially illuminated
9 by a licensing deal for news content.

10 Plaintiffs report that Reuters has resisted compliance with their subpoena request on
11 grounds of overbreadth, vagueness, undue burden, and confidentiality. *Id.* at 6-8. By way of
12 response, Reuters argues that – as a third-party – it is entitled to protection from undue burden and
13 production of confidential information where a substantial need is not demonstrated. *See* Reuters’
14 Opp. (dkt. 348) at 5-7. Reuters also reports that, by way of an offered compromise that would have
15 narrowed the scope of Plaintiffs’ demands, it “offered to produce all executed agreements related
16 to AI training data (which would include the Meta Agreement) and, to the extent they exist,
17 documents and communications with third parties related to all non-executed, proposed, or in
18 pipeline licensing agreements related to AI training data where there was any reasonable chance of
19 aligning on the financial or licensing terms,” but that Plaintiffs’ counsel rejected its offer. *Id.* at 7.
20 Reuters then adds that “Plaintiffs’ counsel [even] declined Reuters’ request that the date for [its]
21 opposition be adjourned to permit the parties to discuss the resolution of the Motion.” *Id.* at 8.

22 In ruling on Plaintiffs’ request to compel this information, the undersigned will note
23 several facts. First, nowhere in Plaintiffs’ motion is there any explanation as to why *the essence* of
24 the information it seeks cannot be garnered either from OpenAI, or through expert witness
25 testimony, or by some other less intrusive means. That is, Plaintiffs have not explained why prying
26 into Reuters’ commercial relationship with Meta is the only way to “establish the existence of a
27 market for copyrighted work as training data for LLMs . . . [such as to] provide evidence of
28 benchmark licensing agreements and market demand . . . [so that Plaintiffs can] calculat[e] actual

1 damages and recovery of lost profits.” *See* Pls. Mot. at 4-5. In their reply brief, Plaintiffs mention
 2 (for the first time, and in a rather conclusory fashion) that they “seek relevant evidence not
 3 otherwise obtainable from OpenAI about the market for textual data² as training data for Large
 4 Language Models.” *See* Pls. Reply (dkt. 367) at 3. Plaintiffs then note that they have sought
 5 licensing deals for AI training data from OpenAI, but they do so without any mention as to the
 6 outcome of having made that request (*i.e.*, without explaining what happened – did OpenAI hand
 7 over a trove of documents? Did OpenAI claim it had nothing? Did OpenAI simply ignore the
 8 request?). *See id.* at 3 n.2.

9 “The determination of substantial need is particularly important in the context of enforcing
 10 a subpoena when discovery of trade secret or confidential commercial information is sought from
 11 non-parties.” *See Gonzales v. Google, Inc.*, 234 F.R.D. 674, 685 (N.D. Cal. 2006) (citing *Mattel*
 12 *Inc. v. Walking Mt. Prods.*, 353 F.3d 792, 814 (9th Cir. 2003). The undersigned will first note that,
 13 in this context, the “substantial need” showing must be concrete, not speculative as is the case
 14 here. *See, e.g., Waymo LLC v. Uber Techs., Inc.*, 2017 U.S. Dist. LEXIS 132721, *10 (N.D. Cal.
 15 Aug. 18, 2017) (“. . . that Uber ‘maybe’ will need discovery from other companies is not even
 16 close to a ruling that Uber has shown a substantial need for compelling the non-parties’ trade
 17 secrets.”). Second, the “substantial need” showing must be just that, substantial. *See, e.g.,*
 18 *Cameron v. Apple Inc. (In re Apple iPhone Antitrust Litig.)*, 2021 U.S. Dist. LEXIS 25194, *12
 19 (N.D. Cal. Jan. 26, 2021) (“In ruling on discovery issues, and in particular to assess ‘substantial
 20 need,’ the Court must sometimes dip its toe into the merits of the case, which inform the
 21 consideration of *how* relevant the requested documents [really] are.”) (emphasis added).

22 Here, the court finds that Plaintiffs have sought what is undeniably confidential
 23 commercial information (if not trade secrets) from one non-party, Reuters, about its relationship
 24 with another non-party, Meta. The court further finds that while Plaintiffs have made a clear case
 25 that OpenAI does not possess the particular documents at issue here – that is, those underlying the
 26 business affairs between Reuters and Meta – what Plaintiffs have not done is to even mention (let
 27

28 ² By using the broad phrase, “textual data,” Plaintiffs gloss over the difference between the news content
 which is the subject of the Reuters/Meta agreement, and the non-news content that is the subject of this case.

1 alone convincingly establish) why the essence of this information (*i.e.*, its upshot or import, which
2 is merely information about the existence of a market for licensing copyrighted “textual data” to
3 train an AI model) cannot be either obtained from OpenAI, or by expert witness testimony, or by
4 any other means that would not involve prying into the confidential commercial information of
5 two non-parties. Moreover, the court finds that the marginal relevance of the information Plaintiffs
6 seek is substantially outweighed by the burden its production would pose to the non-parties whose
7 rights would be implicated by that production. At bottom, the court finds that Plaintiffs have not
8 demonstrated a substantial need for the discovery they seek, and that the requested discovery is not
9 proportional to the needs of the case. Accordingly, Plaintiffs’ Motion (dkt. 314) is **DENIED**.

10 **IT IS SO ORDERED.**

11 Dated: March 5, 2025

12
13 
14 _____
15 ROBERT M. ILLMAN
16 United States Magistrate Judge
17
18
19
20
21
22
23
24
25
26
27
28

United States District Court
Northern District of California