

April 17, 2024
The Honorable Donna M. Ryu
United States District Court for the Northern District of California
Oakland Courthouse, Courtroom 4 – 3rd Floor
1301 Clay Street
Oakland, CA 94612

Re: *Doe I, et al. v. GitHub, Inc., et al.*; Lead Case No. 4:22-cv-06823-JST

Dear Judge Ryu:

Plaintiffs seek an order compelling OpenAI to provide complete responses to Requests for Production Nos. 28-33 and Interrogatory No. 3. OpenAI believes the request should be denied. Counsel met and conferred by video on April 10, 2024, but are at an impasse.¹

A. Request for Production Nos. 28-33

1. Plaintiffs' Position

Request for Production Nos. 28-33 seek documents OpenAI has provided to government entities or political lobbying groups on three discrete topics central to this case: generative AI technology, GitHub, and Copilot. *See* RFP No. 28 (seeking documents OpenAI provided to U.S. regulatory entities); RFP Nos. 31 and 32 (seeking similar submissions to foreign regulators and to Congress); RFP No. 29 (seeking Hart Scott Rodino submissions to the FTC and DOJ regarding acquisitions of interests in OpenAI and GitHub) and RFP No. 30 (seeking SEC filing regarding the same); RFP No. 33 (requesting documents sufficient to identify the lobbyists OpenAI has used to influence generative AI policy in Congress in its favor). In response, OpenAI lodges a series of meritless, boilerplate objections to these RFPs and flat out refuses to provide relevant discovery to which Plaintiffs are entitled.

Communications with government entities are clearly discoverable where they are relevant to issues in a case. *See, e.g., Martin v. Monsanto Co* 2017 WL 5172205, at *5 (C.D. Cal. Apr. 10, 2017) (compelling discovery of communications with state and federal regulatory bodies). The burden of producing such discovery is also minimal since documents like these have already typically been “found, reviewed and organized.” *Waldman v. Wachovia Corp.*, 2009 WL 86763, at *2 (S.D.N.Y. Jan. 12, 2009). OpenAI is also presently valued at 80 billion dollars and can afford to produce these documents. Such discovery requests are also obviously proportional to the needs of a case like this one—historic in every respect, raising issues of heightened public interest and potentially implicating billions of dollars in economic losses. *See Fed. R. Civ. P. 26(b)(1); In re Google RTB Consumer Priv. Litig.*, No. 21CV02155YGRVKD, 2022 WL 1212015, at *1 (N.D. Cal. Apr. 25, 2022).

Contrary to OpenAI’s claims, its statements to government entities about its development of

¹ **Current Case Deadlines:** (1) Fact Discovery Cut-off: September 27, 2024; (2) Expert Discovery Cut-off: February 21, 2025; (3) Dispositive Motion Hearing: May 16, 2024; (4) Class Certification Motion: March 27, 2025; (4) Pretrial Conference and Trial Dates: not scheduled.

Honorable Donna M. Ryu

April 17, 2024

Page 2

generative AI technology, its relationship with Microsoft, and its creation of Copilot is relevant to Plaintiffs' DMCA and contract-based claims. Over the last few years, OpenAI has been the subject of significant government and regulatory scrutiny given its role in developing generative AI. The FTC, for example, recently issued a Civil Investigative Demand to OpenAI seeking, among other things, a detailed description of all the training data OpenAI has used to develop each Large Language Model it has developed, as well as a list of all beneficial owners of 5% or more of OpenAI. Plaintiffs merely seek discovery of documents already provided by OpenAI to regulatory agencies and legislative committees about these issues. The burden of (re)production is negligible as these documents have already been collected and produced. OpenAI's objections as to proportionality and scope are thus unfounded.

These RFPs further seek discovery about OpenAI's ownership and relationship with Microsoft. The FTC, as well as various foreign and domestic regulators, have sought information about this relationship. Documents provided by OpenAI to these entities about its relationship with Microsoft are squarely relevant both to Plaintiffs' claims that Microsoft and OpenAI maintain a joint venture relationship regarding the development of Copilot, as well as to questions of intent. *See, e.g., Rumble, Inc. v. Google LLC*, 2023 WL 3751797, at *7-8 (N.D. Cal. May 31, 2023) (granting plaintiff's motion to compel documents produced to federal and state enforcers and legislators); *In re Plastics Additives Antitrust Litigation*, 2004 WL 2743591, at *12-13 (E.D. Pa. Nov. 29, 2004) (collecting cases).

Relatedly, Plaintiffs' requests for HSR filings OpenAI has provided to federal enforcers and regulators and for SEC filings regarding putative attempts at acquiring interests in OpenAI or GitHub (RFP Nos. 29 & 30) are relevant to Plaintiffs' "joint venture" allegations, as well as Microsoft's influence at OpenAI, for instance, training models without abiding by existing licenses or without otherwise seeking to license training material.

Plaintiffs' Final Proposed Compromise: RFP Nos. 28, 31, 32, and 33, which ask OpenAI to reproduce any regulatory submissions, can be narrowed to documents regarding LLMs, GitHub, or Copilot, and the ownership and control of OpenAI. OpenAI should produce responsive documents to RFPs 29 and 30.

2. OpenAI's Position

RFP Nos. 28, 31, 32, and 33 seek documents that are irrelevant. Only two claims remain against OpenAI (with a motion to dismiss pending)—one for violation of 17 U.S.C. § 1202(b) (the "DMCA claim"), and another for breach of the licenses that allegedly apply to Plaintiffs' code. Both claims relate only to one OpenAI product: Codex. These claims will depend on a discrete set of facts—the contents of Plaintiffs' code, the alleged copyright management information (CMI) conveyed in connection with that code, and how OpenAI allegedly "trained" the models used for Codex as to that CMI. But Plaintiffs' RFPs seek a practically unbounded universe of documents, so long as they were given to a government agency. Because OpenAI is an industry leader, it has produced documents to many government entities on a wide variety of topics relating to generative AI. For example, OpenAI is producing documents to the FTC in connection with an investigation focused on privacy and consumer protection; that investigation has nothing to do with Plaintiffs' claims, but Plaintiffs demand the documents anyway.

Honorable Donna M. Ryu

April 17, 2024

Page 3

Courts have repeatedly recognized that requesting “all documents produced in another matter is not generally proper” because the propounding party “cannot meet its burden to establish relevance, as the propounding party is not in a position to even know what they are actually asking for.” *Ludlow v. Flowers Foods, Inc.*, No. 18-cv-1190, 2019 WL 6252926, at *18 (S.D. Cal. Nov. 22, 2019); *see also King Cnty. v. Merrill Lynch & Co., Inc.*, No. C10-1156-RSM, 2011 WL 3438491, at *3 (W.D. Wash. Aug. 5, 2011) (“Cloned discovery, requesting all documents produced or received during other litigation or investigations, is irrelevant and immaterial unless the fact that particular documents were produced or received by a party is relevant to the subject matter of the subject case.”) (cleaned up); *Wilmington Trust Co. v. Boeing Co.*, No. C20-0402-RSM-MAT, 2020 WL 4125106, at *2 (W.D. Wash. July 20, 2020) (rejecting discovery of “all documents produced or relied upon” in “responding to requests from governmental agencies”); *N.J. Carpenters Health Fund v. DLJ Mortg. Cap., Inc.*, No. 08 Civ. 5653 (PAC), 2012 WL 13135408 (S.D.N.Y. Mar. 2, 2012) (denying discovery request that was described as “you gave some documents to the government concerning another investigation, so give them to me”).

The cases that Plaintiffs cite do not help them. *Martin* involved false advertising claims based on Monsanto’s using the phrase “makes up to.” 2017 WL5172205 at *1. The plaintiff sought documents produced to the government concerning the use of that phrase. *Id.* at *3. The court overruled Monsanto’s relevance objection because a “regulatory body’s interpretation of a Monsanto claim or statement similar to ‘makes up to’ may implicate consumer reliance and purchasing behavior for its products.” *Id.* at *5. Plaintiffs’ other cases, *Rumble* and *In re Plastics Additives Antitrust Litigation*, involved antitrust plaintiffs seeking documents produced to the government in related antitrust investigations. But here, Plaintiffs have “fail[ed] to establish a clear logical link between these RFPs and any claim or defense in this case.” *Elgindy v. Aga Serv. Co.*, No. 20-cv-06304, 2021 WL 5083761, at *4 (N.D. Cal. Nov. 2, 2021). Plaintiffs nevertheless argue that they are entitled to documents based on relevance to their “allegations.” But that is not the rule. Plaintiffs must show relevance to an element of a *claim* or *defense*. *Elgindy*, 2021 WL 5083761 at *2; *see also* Fed. R. Civ. P. 26(b). They haven’t done so.

RFP Nos. 29 and 30 are equally problematic. Documents about ownership interests in OpenAI provided to government agencies are not relevant to CMI, Plaintiffs’ alleged licenses or code, or anything else that would “tend[] to make the existence of any fact that is of consequence to the determination of the claim[s] more or less probable.” *See Manzo v. Cnty. of Santa Clara*, No. 17-cv-01099, 2019 WL 2866047, at *2 (N.D. Cal. July 3, 2019); *see also Blast Mot., Inc. v. Zepp Labs., Inc.*, 2015 WL 12564315, at *3 (S.D. Cal. Dec. 2, 2015) (denying motion to compel documents sufficient to identify investors with greater than 5% ownership interest as irrelevant to patent claims); *AngioScore, Inc. v. TriReme Med., Inc.*, No. 12-cv-03393, 2014 WL 6706873, at *2-3 (N.D. Cal. Nov. 25, 2014) (ownership interests of board of directors not relevant to patent infringement and state law claims); *Sadofsky v. Fiesta Prods., LLC*, 252 F.R.D. 143, 153 (E.D.N.Y. 2008) (ownership interests in defendant not relevant to dispute over patent royalty agreement). As with Request Nos. 28, 31, 32, and 33, these Requests also seek impermissibly broad and overly burdensome “cloned discovery.”

Additionally, the Requests impose an undue burden. Requests for “duplicate searches”—one for relevant documents generally and another for relevant documents provided to the government in

Honorable Donna M. Ryu

April 17, 2024

Page 4

unrelated matters—are “definitionally unduly burdensome, as it would consume resources without providing any additional benefit to the propounding party.” *Ludlow*, 2019 WL 6252926, at *18; *see also In re ATM Fee Antitrust Litig.*, No. 04-cv-02676, 2007 WL 1827635, at *5 (N.D. Cal. June 25, 2007) (denying motion to compel documents “submitted pursuant to the Hart-Scott-Rodino Act” because “the potential for turning up a needle does not require Defendants to sift through this haystack”). Plaintiffs misunderstand *Waldman*. That case addressed whether a plaintiff had shown “undue prejudice” to justify lifting an automatic discovery stay under the Private Securities Litigation Reform Act, an entirely different standard. In any event, the court there found that the burden of lifting such a stay would be slight *if* “a defendant has already found, reviewed and organized the documents.” 2009 WL 86763 at *2 (cleaned up). That is not the case here. Here, OpenAI would need to re-review any previously produced documents to determine whether they are relevant and unobjectionable in *this* case. Since OpenAI is already searching for documents in response to Plaintiffs’ properly propounded RFPs, doing so would be the sort of “definitionally unduly burdensome” “duplicate search[]” that courts have rejected. *Ludlow*, 2019 WL 6252926, at *18.

OpenAI’s Final Proposed Compromise: Plaintiffs should serve discovery requests targeted at documents relevant and proportional to the claims in this case.

B. Interrogatory No. 3

1. Plaintiffs’ Position

Interrogatory No. 3 seeks information about individuals and entities with a beneficial ownership interest in OpenAI of greater than 5%. Ex. 1. Beneficial ownership of greater than 5% presumptively confers a material level of control over a corporation. The Court should order OpenAI to respond to Interrogatory No. 3.

Rule 26 proportionality factors apply with equal force here. The interrogatory seeks identification of persons or entities who exercise presumptive influence over OpenAI’s management and direction, including decisions relevant to this case. This interrogatory follows long established standards for determining influence and control based on beneficial ownership.

For example, sections 13(d) and 13(g) of the Securities Exchange Act generally require any person(s) who acquire 5% or more of an issuer’s Section 13(d) securities to report such ownership to the SEC in order “to alert investors to potential changes in control[.]” *Lowinger v. Morgan Stanley & Co. LLC*, 841 F.3d 122, 132 (2d Cir. 2016). In addition, the F.C.C. determined 5% is an appropriate ownership benchmark for the purposes of its efforts to limit the multiple ownership of broadcast facilities because an owner of 5% or more “would ordinarily be one of the two or three largest shareholders,” large enough “to be in a position to command the attention of management.” *Attribution of Ownership Interests*, 97 FCC 2d 997, 1005-08 (1984). This reasoning was upheld in a subsequent Court of Appeals decision. *Time Warner Ent. Co., L.P. v. F.C.C.*, 240 F.3d 1126, 1140 (D.C. Cir. 2001). Such individuals are therefore likely to possess discoverable information relevant to this action.

Honorable Donna M. Ryu

April 17, 2024

Page 5

Interrogatory No. 3 simply asks Defendant to *identify* the names of individuals or entities that own an interest of 5% or more. This is a routine first-wave discovery request. Such individuals or entities may have discoverable documents or information relevant to the claims in this action. Other such individuals may have exerted control or voiced concerns over issues relevant in this case (like the unlawful use of training data without regard to licenses). *See Time Warner*, 240 F.3d at 1140. There is no burden of producing it. To the extent OpenAI has privacy concerns, the Protective Order can adequately address them.

Plaintiffs' Final Proposed Compromise: OpenAI should supplement its response to Interrogatory 3 and identify all beneficial owners of 5% or more of OpenAI.

2. OpenAI's Position

Plaintiffs claim that OpenAI's investors may have relevant documents or control over OpenAI's decisions. Neither argument shows relevance. In *Blast Motion, Inc. v. Zepp Labs, Inc.*, the court rejected a motion to compel production in response to a substantially identical RFP seeking "documents sufficient to identify any investor with an ownership interest of at least 5%" in the defendant company. 2015 WL 12564315, at *3. Like here, plaintiff in that case sought investor information to "help it identify additional potential witnesses" and to "assist [it] in understanding who has a financial interest or stake in the case." *Id.* But the court denied the motion, explaining that plaintiff's "desire to learn who has a financial interest in this litigation is not relevant" and that plaintiff "does not need this information to identify potential witnesses." *Id.* The same is true here. Plaintiffs' "desire to learn" who has "control" over OpenAI is overly broad on its face and untethered to any element of the remaining DMCA or breach of contract claims.

The cases that Plaintiffs cite do not suggest that investors, as a class, necessarily have relevant information or documents. Plaintiffs show no relevant connection between the SEC's or FCC's views on holding a 5% interest in a company and the claims in this case. Because Plaintiffs have not met their burden to establish the relevance of the information sought by its request, the Court should deny Plaintiffs' motion. *Elgindy*, 2021 WL 5083761, at *2.

OpenAI's Final Proposed Compromise: This interrogatory seeks irrelevant information and OpenAI will not respond to it.

FILER'S ATTESTATION

Pursuant to Civil Local Rule 5-1(h)(3), I attest under penalty of perjury that concurrence in the filing of the foregoing document has been obtained from all other signatories to this document.

Dated: April 18, 2024

By: /s/ Joseph R. Saveri
Joseph R. Saveri