

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

In Re Mosaic LLM Litigation

Case No. [24-cv-01451-CRB](#) (LJC)

**ORDER REGARDING PLAINTIFFS’
SUBPOENA TO MICROSOFT**

Re: Dkt. No. 139

Pending before the Court is Plaintiffs’ and non-party Microsoft Corporation’s (Microsoft’s) joint discovery letter regarding a subpoena Plaintiffs served on Microsoft. ECF No. 139. The dispute was heard on July 31, 2025. ECF No. 149.

Plaintiffs served a subpoena on Microsoft seeking, at Request for Production No. 1 (RFP No. 1), all licensing agreements Microsoft entered into for AI training data and, at Request for Production No. 2 (RFP No. 2), all documents and communications “related to licensing books for use as AI training data.” ECF No. 139-1 at 2-3. Microsoft produced all executed licensing agreements it has entered into “that provide for Microsoft to receive, access, and/or use text data in training LLMs[,]” but objects to producing other “documents and communications.” ECF No. 139 at 3. It argues that such documents are not relevant to the case and production would be overly burdensome. *Id.* Plaintiffs contend that the documents are highly relevant such that any burden on Microsoft is warranted. Having considered the record in this case, as well as the parties’ briefing and oral arguments, for the foregoing reasons, the Court GRANTS IN PART Plaintiffs’ request to compel Microsoft to produce documents in response to RFP No. 2 by narrowing the scope of the subpoena.

I. LEGAL STANDARD

Rule 45 of the Federal Rules of Civil Procedure permits the use of a subpoena to command a nonparty to produce designated documents and electronically stored information. Fed. R. Civ. P. 45(a)(1)(A)(iii). “The scope of discovery under Rule 45 is the same as under Rule 26(b).” *Waymo LLC v. Uber Techs., Inc.*, No. 17-cv-00939, 2017 WL 2929439, at *2 (N.D. Cal. July 7, 2017); *see also Maplebear Inc. v. Uber Techs., Inc.*, No. 21-mc-80007, 2021 WL 1845535, at *1 (N.D. Cal. Mar. 23, 2021) (citing Fed. R. Civ. P. 45 Advisory Comm.’s Note (1970)). Federal Rule of Civil Procedure 26(b)(1) provides that parties to a lawsuit “may obtain discovery regarding any nonprivileged matter that is relevant to any party’s claim[s] or defense[s] and proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy, the parties’ relative access to relevant information, the parties’ resources, and the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit.”

“A party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena.” Fed. R. Civ. P. 45(d)(1). Courts are required to “enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney’s fees—on a party or attorney who fails to comply.” *Id.* An order quashing or modifying a subpoena is required where a subpoena, among other things, “requires disclosure of privileged or other protected matter, if no exception or waiver applies” or “subjects a person to undue burden.” Fed. R. Civ. P. 45(d)(3)(A). Courts may, on motion, quash or modify a subpoena that requires the disclosure of a “trade secret or other confidential research, development, or commercial information[.]” Fed. R. Civ. P. 45(d)(3)(B)(i). Alternatively, when such a disclosure is required by a subpoena, “the court may, instead of quashing or modifying a subpoena, order . . . production under specified conditions if the serving party” shows “substantial need for the...material that cannot be otherwise met without undue hardship[.]” and “ensures that the subpoenaed person will be reasonably compensated.” Fed. R. Civ. P. 45(d)(3)(C).

The “Ninth Circuit has long held that nonparties subject to discovery requests deserve

extra protection from the courts,” and these protections underlie Rule 45’s protections. *Maplebear Inc. v. Uber Techs., Inc.*, No. 21-mc-80007, 2021 WL 1845535, at *1 (N.D. Cal. Mar. 23, 2021) (quotations omitted); *Gonzales v. Google, Inc.*, 234 F.R.D 674, 685 (N.D. Cal. 2006); see *Dart Industries Co., Inc. v. Westwood Chemical Co., Inc.*, 649 F.2d 646, 649 (9th Cir. 1980) (“While discovery is a valuable right and should not be unnecessarily restricted... the necessary restriction may be broader when a nonparty is the target of discovery.”) (quotations and citation omitted) (pre-dating the 1991 amendment to Rule 45 that added subdivision (c), the precursor to subdivision (d), stating the rights of non-party witnesses). This is because “[n]onparty witnesses are powerless to control the scope of litigation and discovery[.]” *U.S. v. Columbia Broad. Sys., Inc.*, 666 F.2d 364, 371 (9th Cir. 1982).

“A nonparty commanded to produce documents pursuant to a Rule 45 subpoena may challenge the subpoena in one of three ways: (1) by written objection; (2) by moving to quash or modify the subpoena; or (3) by moving for a protective order.” *Sacks Holdings, Inc. v. Vaidya*, No. 24-mc-80197, 2024 WL 4730424, at *3 (N.D. Cal. Nov. 7, 2024); see Fed. R. Civ. P. 26(c); Fed. R. Civ. P. 45(d)(2)(B), (d)(3). If a nonparty objects to a subpoena, the serving party may request that the Court order compliance. Fed R. Civ. P. 45(d)(2)(B)(i). The party seeking discovery bears the burden of establishing relevance, and the objecting party bears “the burden of persuasion to show that the discovery should not be allowed.” *Sacks Holdings*, 2024 WL 4730424, at *3.

II. RELEVANCE

Plaintiffs advance two arguments for why documents responsive to RFP No. 2 are relevant. They argue that “evidence of the licensing market is relevant to factor four of the fair use defense that Defendants have raised” and that evidence of “the existence and continued development of a licensing market for training data” is relevant to damages. ECF No. 139 at 2. Defendants contend that the executed licenses they have produced are more than sufficient and that the potential market for licensing training data is irrelevant to the fair use analysis. The Court addresses Plaintiffs’ arguments in turn.

A. Fair Use

Plaintiffs argue that the market for licensing written work for AI training data is relevant to Defendants' fair use defense. The fair use doctrine establishes that "the fair use of a copyrighted work...for purposes such as criticism, comment, news reporting, teaching..., scholarship, or research, is not an infringement of copyright." *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417, 2025 WL 1752484, at *3 (N.D. Cal. June 25, 2025) (citing 17 U.S.C. § 107). While fair use is a "flexible concept," the Copyright Act lists four factors courts consider to determine whether use of a copyrighted work is fair: "(1) the purpose and character of the use...; (2) the nature of the copyrighted work; (3) the amount and sustainability of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work." *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 527 (2023); 17 U.S.C. § 107. Plaintiffs contend that Microsoft's communications and negotiations regarding licensing books for generative AI training purposes is relevant to the fourth factor, as the existence and nature of the market for licensing works for AI training may support Plaintiffs' argument that Defendants' alleged infringement harmed the market for licensing Plaintiffs' copyrighted works.

The heart of the issue is if Plaintiffs' theory of market harm is cognizable under the fair use analysis. Factor four addresses the risk of "harm of market substitution": whether the defendant's allegedly infringing conduct or "unrestricted and widespread conduct of the sort engaged in by the defendant...would result in a substantially adverse impact on the potential market for the original." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590, 593 (1994) (quotations omitted). When an alleged infringer's derivative use "amounts to mere duplication of the entirety of an original, it clearly supersedes the objects...of the original and serves as a market replacement for it, making it likely that cognizable market harm to the original will occur...But when, on the contrary, the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred." *Id.* at 591 (citations and quotations omitted). Plaintiffs are authors who hold copyrights in their written work. Copying their books without permission, reproducing them in large part or in their entirety, and selling the copies on the cheap would

presumably run afoul of factor four, as consumers could buy the unauthorized copies rather than Plaintiffs' works. *See Campbell*, 510 U.S. at 590. Arguably, so would copying their books and producing versions "similar enough" in content or style to the originals that readers abandoned the original books in favor of the derivatives. *Kadrey*, 2025 WL 1752484, at *15. In their discovery letter, Plaintiffs do not argue that they are being—or have the potential to be—harmed by readers switching from buying and reading their books to buying and reading content generated by Defendants. *See* ECF No. 139. Rather, they contend that they are harmed by Defendants' alleged infringement because it adversely impacts the market for licensing their copyrighted works for the purpose of AI training. *Id.* at 2.

This raises a circularity issue. *See Google LLC v. Oracle America, Inc.*, 593 U.S. 1, 38 (2021). If the "potential market" is defined as the market for licensing a copyrighted work for the "very use" employed by the alleged infringer, then factor four would always weigh in favor of the copyright holder. *Tresona Multimedia, LLC v. Burbank High Sch. Vocal Music Ass'n*, 953 F.3d 638, 652 (9th Cir. 2020) (quotations omitted). Two recent decisions in this district that have addressed this issue in the context of generative AI cases have found at summary judgment that "to prevent the fourth factor analysis from becoming circular and favoring the rightsholder in every case, harm from the loss of fees paid to license a work for a transformative purpose is not cognizable." *Kadrey*, 2025 WL 1752484, at *16; *Bartz v. Anthropic PBC*, No. C 24-05417, 2025 WL 1741691, at *17 (N.D. Cal. June 23, 2025) ("[Plaintiff] Authors next contend that training LLMs displaced (or will) an emerging market for licensing their works for the narrow purpose of training LLMs... such a market for that use is not one the Copyright Act entitles Authors to exploit.").¹

Microsoft urges the Court to follow *Kadrey* and *Bartz* and find that the negative impact on Plaintiffs' ability to license their work for generative AI training is not a cognizable form of market harm under factor four, and thus not relevant as a matter of law. For the following reasons,

¹ Although Judge Illman addressed the issue of Microsoft producing documents and communication regarding licensing, finding that such production was not warranted, his order did not address the relevance of such documents to the defendants' fair use defense. *Tremblay v. OpenAI, Inc.*, No. 23-cv-03223, 2025 WL 729674, at *1 (N.D. Cal. Mar. 6, 2025).

the Court declines to do so. First, this issue is being raised in a discovery letter brief. The undersigned must determine if documents Plaintiffs seek are “relevant to any party’s claim or defense[.]” Fed. R. Civ. P. 26(b). “In ruling on discovery issues, and in particular to assess ‘substantial need,’ the Court must sometimes dip its toe into the merits of the case, which inform the consideration of how relevant the requested documents are.” *In re Apple iPhone Antitrust Litig.*, No. 11-cv-06714, 2021 WL 485709, at *3 (N.D. Cal. Jan. 26, 2021). At the same time, for purposes of discovery, relevance is defined broadly. *Snipes v. United States*, 334 F.R.D. 548, 550 (N.D. Cal. 2020). “Relevant information for purposes of discovery is information ‘reasonably calculated to lead to the discovery of admissible evidence.’” *Survivor Media, Inc. v. Survivor Prods.*, 406 F.3d 625, 635 (9th Cir. 2005) (citing Fed. R. Civ. P. 26(b)). “This requirement is liberally construed to permit the discovery of information which ultimately may not be admissible at trial.” *Gonzales*, 234 F.R.D at 680.

Second, while the conclusions in *Kadrey* and *Bartz* are well-reasoned, the Ninth Circuit has not yet addressed factor four in the context of generative AI. This remains an unsettled area of law. For example, although Judge Alsup found that factor four weighed in favor of fair use, he differentiated between the types of use that infringed on “something the Copyright Act properly protected” versus those “for which a copyright owner cannot rightly expect to control.” *Bartz*, 2025 WL 1741691, at *17. As generative AI proliferates, the right to license one’s creative work to train generative AI models may arguably become—or may already be—something copyright owners “rightly expect to control.” *Id.*; see *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 568 (1985) (citing senate report stating that “a use that supplants any part of the *normal market for a copyrighted work* would ordinarily be considered an infringement”) (emphasis added). Five years ago, a market for licensing one’s books for AI training could hardly be considered “normal” or “reasonable[.]” See *Harper & Row*, 471 U.S. at 586; *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 930 (2d Cir. 1994). Five years from now, it may be as “normal” a market as the market for licensing books for movies or television shows. Given the rapid evolution of generative AI and entrepreneurial responses to this emerging technology, what “the normal market for a copyrighted work” is in flux. *Harper & Row*, 471 U.S. at 586.

Documents regarding Microsoft’s negotiations with third parties to license books for training AI models may show that such a market is “normal[.]” “reasonable, or likely to be developed” and thus such a market could be cognizable under factor four. 4 Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 13F.08 (2025); *Texaco Inc.*, 60 F.3d at 930. And, as Plaintiffs note, the Second Circuit has held that the existence of a “ready market or means to pay for the use” of copyrighted work can, under factor four, make unauthorized use “less fair.” *Texaco*, 60 F.3d at 931. Microsoft’s negotiations regarding licensing with publishers and other rightsholders are relevant to support or belie the existence of a “ready market[.]” *Id.*

The undersigned is not aware of controlling authority that forecloses a finding that evidence of an existing market for licensing material for training AI bears on the factor four analysis. And, if Judge Breyer considers the market for licensing for AI-training purposes under factor four, documents responsive to RFP No. 2 would be highly important to resolving the fair use issue, which may very well be dispositive of Plaintiffs’ claims. *See* Fed. R. Civ. P. 26(b) (permitting discovery that is “relevant to any party’s claim or defense and proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy...[and] the importance of the discovery in resolving the issues”). The door has not yet been shut on Plaintiffs’ arguments regarding the market for licensing and, as long as this theory is active, documents related to the licensing market are substantially needed and highly relevant. *Cf. In re Apple iPhone* Litigation, 2021 WL 485709, at *3 (in the context of antitrust law, finding “Apple’s definition of the relevant market, which is the linchpin that would make the requested documents relevant at all” inconsistent with an order by the presiding judge and thus determining that the “requested documents would therefore seem to have very limited relevance”). Finding that documents relating to the licensing market are not relevant at this point in litigation would prematurely foreclose Plaintiffs’ ability to advance their claim and shift what is likely a central issue in this case from resolution at summary judgment to a definitive resolution through a discovery dispute. The Court declines Microsoft’s request to issue such a ruling in the absence of controlling authority, particularly as the fair use analysis ought not be “simplified with bright-line rules” and requires fact-specific, “case-by-case analysis.” *Campbell*, 510 U.S. at 578.

Moreover, the Supreme Court has instructed that under factor four, courts “must take into account the public benefits the copying will likely produce” relative to the “losses to copyright owner[.]” *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 35 (2021) (citing *MCA, INC. v. Wilson*, 677 F. 2d 180, 182 (2d Cir. 1981)). While this analysis is not separate from circularity issue—courts must take “into account as well the nature of the source of the loss” which the undersigned understands to mean if the loss is of the type protected by the Copyright Act—the undersigned recognizes that the existence of a licensing market may undercut the public benefit of infringement. That is, accepting but not deciding that there is a “public benefit” from AI companies developing effective language-generative models trained on large volumes of written text, the fact that some companies have obtained licenses for written work to train their models suggests that the public can receive the same benefit without the alleged infringement. *Id.*

The Court agrees with Plaintiffs that information beyond just the executed agreements that Microsoft has produced is relevant to assessing the licensing market for training generative AI models. The executed agreements show the end results. The communications—both those leading to executed agreements and those where no agreement was reached—shed light on the nature of the market: what terms and prices parties initially proposed, negotiations regarding valuation, alternative licensing structures that were proposed and rejected, and why Microsoft or other parties ended negotiations. The Court accordingly finds that documents responsive to RFP No. 2 are relevant and Plaintiffs have shown they are necessary to advance their fair use argument.

B. Damages

Plaintiffs also contend that “negotiations or valuations Microsoft engaged in with publishers or copyright holders implicate potential class works will likely help calculate damages in this case.” ECF No. 139 at 2. Copyright holders are entitled to recover the “actual damages suffered by him or her as a result of the infringement, and any profits of the infringer that are attributable to the infringement” or statutory damages. 17 U.S.C. § 504. One measure of actual damages is the fair market value of a hypothetical license. *Oracle Corp. v. SAP AG*, 765 F.3d 1081, 1091 (9th Cir. 2014); *Jarvis v. K2 Inc.*, 486 F.3d 526, 533 (9th Cir. 2007) (“[I]n situations where the infringer could have bargained with the copyright owner to purchase the right to use the

work, actual damages are what a willing buyer would have been reasonably required to pay to a willing seller for plaintiffs’ work.”) (quotations omitted). “The touchstone for hypothetical-license damages is ‘the range of [the license’s] reasonable market value.’” *Oracle Corp.*, 765 F.3d at 1088 (quoting *Polar Bear Prods., Inc. v. Timex Corp.*, 384 F. 3d 700, 709 (9th Cir. 2004)). “An award of hypothetical-license damages is appropriate ‘provided the amount is not based on undue speculation.’” *Id.* (citing *Polar Bear*, 384 F. 3d at 709). Evidence that a copyright holder “would have reached a licensing agreement with the infringer” or of “benchmark licenses in the industry approximating the hypothetical license” may help the rightsholder “establish the amount of such damages without undue speculation[.]” *Id.* at 1093.

Although no named Plaintiffs have licensed their copyrighted books to Microsoft for purposes of AI training, Microsoft’s licenses with other rightsholders—including potential class members—are relevant to show the “benchmark licenses in the industry.” *Williams v. Gaye*, 895 F.3d 1106, 1129 (9th Cir. 2018); *Moebius v. HB USA Holdings, Inc.*, No. CV 12109, 2022 WL 17219089, at *7 (C.D. Cal. May 25, 2022) (collecting cases). But Microsoft has already provided Plaintiffs with their executed licenses. While the Court recognizes that evidence of Microsoft’s “negotiations or valuations” regarding licensing could provide context regarding the industry, Plaintiffs have not shown that they have a substantial need for this information in addition to the already-produced licenses.

III. BURDEN

“To lessen the burden on Microsoft,” Plaintiffs offered to limit the scope of their request to Microsoft’s “documents that it has already produced in the ongoing *In re OpenAI*” MDL, to which Microsoft is a party. ECF No. 139 at 2. Microsoft contends that this “compromise” does not “sufficiently relieve” the burden of production because 1) they will still have to search through the over two hundred thousand documents they produced in the MDL to find documents responsive to RFP No. 2, and 2) production would require them to produce confidential business information and communications with third parties unrelated to this litigation. *Id.* at 5. Microsoft also argued at the hearing that Plaintiffs’ request for all documents and communications related to licensing is vague and overbroad, such that it is difficult to comply with. RFP No. 2 seeks “All Documents

1 and Communications related to licensing books for use as AI training data.” ECF No. 139-1 at 2.
2 This includes all communications and negotiations with third parties regarding potential licensing
3 agreements, without cabining production to advanced negotiations. Microsoft noted that it had
4 entered into licenses with fourteen to fifteen rightsholders, but, without clarity of what level of
5 communications constituted “formal negotiations[,]” could not provide an estimate of how many
6 others it had negotiated with.

7 The undersigned is cognizant that Microsoft, as a nonparty, is “afforded additional
8 protection against” overbroad discovery requests. ECF No. 139 at 4; *see Maplebear*, 2021 WL
9 1845535, at *1. And it credits Microsoft’s argument that it cannot just pull responsive documents
10 off the shelf but would have to “conduct new searches” to locate and review documents responsive
11 to RFP No. 2 among the nearly two hundred thousand documents produced in the MDL. ECF No.
12 139 at 5. This situation is dissimilar to *In re McKesson Governmental Entities Average Wholesale*
13 *Price Litigation*, where no re-review was warranted. *See* 264 F.R.D. 595 (N.D. Cal. 2009).
14 Moreover, “[e]very time a business is required to permit its confidential information to be shared
15 there is a risk, and thus a burden, that it will not be maintained as confidential notwithstanding
16 protective orders.” *Pac. Wine Distributors, Inc. v. Vitol Inc.*, No. 20-cv-03131, 2022 WL
17 1489474, at *2 (N.D. Cal. May 11, 2022). The undersigned must balance Plaintiffs’ need for the
18 requested information with the injury that Microsoft contends will result from disclosure.
19 *Gonzales*, 234 F.R.D. at 685. Here, Microsoft would disclose not just its own business
20 information but, presumably, the business information of the third parties it negotiated with.
21 Although there is a protective order entered in this case, and Plaintiffs appear willing to designate
22 documents produced in response to RFP No. 2 as confidential or highly confidential, as Judge
23 Corley noted in *Vitol*, protective orders are not failsafe. *See* 2022 WL 1489474, at *2. At the
24 same time, as the Court has determined that RFP No. 2 seeks documents that are relevant to a
25 central claim of the case. *See* Fed. R. Civ. P. 26(b).

26 In light of these concerns, the undersigned finds that RFP No. 2 is overbroad and places an
27 undue burden on nonparty Microsoft, even with Plaintiffs’ offer to limit it to documents Microsoft
28 produced in the *In re OpenAI* MDL. *See Convolv*, 2011 WL 1766486, at *2 (“Requests to non-

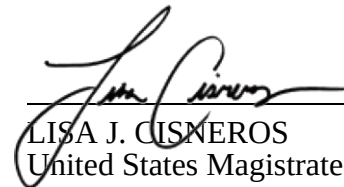
1 parties should be narrowly drawn to meet specific needs for information.”). The Court declines to
2 order Microsoft to produce all documents and communications “related to licensing books for use
3 as AI training data.” ECF No. 139-1 at 2. Instead, the undersigned modifies the subpoena to limit
4 Microsoft’s production to its communications with and documents regarding the entities with
5 which it entered into licensing agreements for AI Training Data, where the communications or
6 documents indicate potential or proposed terms for licensing. Microsoft and Plaintiffs are directed
7 to designate documents produced in response to RFP No. 2 as “Highly Confidential – Attorneys’
8 Eyes Only.”

9 **IV. CONCLUSION**

10 For the forgoing reasons, Plaintiffs’ request that Microsoft be compelled respond to RFP
11 No. 2 is GRANTED IN PART.

12
13 **IT IS SO ORDERED.**

14 Dated: July 8, 2025

15
16
17 
18 LISA J. CISNEROS
19 United States Magistrate Judge
20
21
22
23
24
25
26
27
28