United States District Court
Northern District of California

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

CONCORD MUSIC GROUP, INC., et al., Plaintiffs, v. ANTHROPIC PBC, Defendant.	Case No. 24-cv-03811-EKL (SVK) **ORDER ON JOINT DISCOVERY SUBMISSIONS** Re: Dkt. Nos. 328, 341, 345

Before the Court are the Parties' Joint Discovery Submissions regarding three discovery disputes. First, Defendant Anthropic PBC ("Anthropic") seeks to compel Plaintiffs ("Publishers") to disclose and produce currently undisclosed prompts to and outputs from Claude (and related settings) that Publishers and their agents submitted and obtained in the course of their legal investigation. Dkt. 328 at 2 (relating to Interrogatory no. 1 and Requests for Production ("RFP") nos. 37-43). Second, the Parties disagree as to the appropriate scope of the sampling protocol in response to this Court's prior order. Dkt. 341; *see also* Dkt. 318. Third, Publishers challenge certain confidentiality designations made by Anthropic as overbroad and improper. Dkt. 345. These matters came on for hearing on May 13, 2025. Having considered the Parties' arguments, the relevant law and the record in this action, the Court **DENIES without prejudice** Anthropic's request to compel production of Publishers' prompts, outputs, and settings; **ORDERS production of a sample of 5 million prompt-output pairs** according to the protocol set forth below; and **SUSTAINS-IN-PART** and **OVERRULES-IN-PART** Publishers' challenge to Anthropic's confidentiality designations.

////

////

////

I. DKT. NO. 328: ANTHROPIC'S MOTION TO COMPEL PRODUCTION OF PUBLISHERS' UNDISCLOSED PROMPTS AND OUTPUTS AND THE SETTINGS USED THEREFORE

A. Legal Standard

Federal Rule of Civil Procedure 26 protects disclosure of documents and tangible things "prepared in anticipation of litigation or for trial." Fed. R. Civ. P. 26(b)(3)(A); *see In re Grand Jury Subpoena (Mark Torf/Torf Env't Mgmt.)*, 357 F.3d 900, 906 (9th Cir. 2004). This includes attorney work product, which may be either fact work product or opinion work product. "[O]pinion work product" includes "an attorney's mental impressions, conclusions, opinions, or legal theories developed in anticipation of litigation." *Republic of Ecuador v. Mackay*, 742 F.3d 860, 869 n.3 (9th Cir. 2014); *see* Fed. R. Civ. P. 26(b)(3)(B). It "is virtually undiscoverable." *Id.*

However, "[t]he privilege derived from the work-product doctrine is not absolute. Like other qualified privileges, it may be waived." *United States v. Sanmina Corp.*, 968 F.3d 1107, 1119 (9th Cir. 2020). "Similar to the waiver of the attorney-client privilege, a litigant can waive work-product protection to the extent that he reveals or places the work product at issue during the course of litigation." *Id.* In other words, under the "fairness principle," a party cannot "us[e] the privilege as both a shield and a sword." *Id.* at 1117 (explaining the fairness principle in context of attorney-client privilege).

B. Discussion

In this case, Publishers conducted a pre-suit investigation into potential infringement by Anthropic and relied upon certain prompts and outputs from their investigation in support of their allegations of infringement. *See* Dkt. 1; Dkt. 337-2 (Ex. B); Dkt. 369 at 1-2. Publishers represent that they have produced all prompts and outputs on which they have relied, totaling nearly 5,000 prompt-output pairs. Dkt. 369 at 2-3. Anthropic has now served broad discovery requests directed to *all* prompts and outputs including those not relied upon by Publishers – in other words, those that presumably did not support claims of infringement. *See* Dkts. 328-1–2 (Interrogatory no. 1 and RFP nos. 37-43). In the joint statement, Anthropic argues that either (a) the unrelied-upon prompts/outputs are not privileged or (b), if privileged, that the privilege has been waived. Dkt. 328 at 1-5. On the record before it, the Court disagrees.

Anthropic's initial argument, that the information it seeks (undisclosed prompts and outputs, and the settings therefore) is not privileged is unpersuasive. Publishers cite cases where courts, including in this District, have found precisely this information to constitute attorney work product. *E.g.*, *Tremblay v. OpenAI, Inc.*, No. 23-cv-03223-AMO, 2024 WL 3748003, at *2-*3 (N.D. Cal. Aug. 8, 2024) (explaining that the underlying magistrate judge's order had found "that the account settings and negative test results [were] fact work product [but] that Plaintiffs waived the ability to assert work product" and holding that this was a misapplication of law because "the ChatGPT prompts were queries crafted by counsel and contain counsel's mental impressions and opinions about how to interrogate ChatGPT" and were thus opinion, not fact, work product). Anthropic distinguishes only *Tremblay*'s denial of waiver, based on the plaintiffs' use of outputs there as being unlike this case, but does not distinguish the basic finding that the failed prompts and related settings are attorney work product. *See* Dkt. 328 at 5. This Court agrees with *Tremblay* in that regard.

The closer issue is the extent to which Publishers have waived the attorney work product protection. Publishers admit that they have relied on certain prompts and outputs in their First Amended Complaint (Dkt. 337) and in various other filings, including their original and renewed motions for preliminary injunction and supporting declarations but also represent that they have produced all prompts and outputs relied upon. *See* Dkt. 369 ("Publishers have produced all Claude prompts and outputs from their investigation on which they have relied upon to date. In total, Publishers have produced approximately 4,659 Claude prompt/output records…."); *see also* Dkt. 372 ("Hrg. Tr.") at 50:16-51:18. Under the "fairness principle" or sword-and-shield doctrine, there has been at least a limited waiver here; indeed, Publishers have produced nearly 5,000 prompt-output pairs upon which they rely. The issue is how far the waiver reaches.

The Ninth Circuit has made clear that "the scope of [a work product] waiver must be 'closely tailored ... to the needs of the opposing party' and limited to what is necessary to rectify any unfair advantage gained…." *Sanmina*, 968 F.3d at 1124. Anthropic claims waiver by pointing to Publishers' "ease of use" or "massive use" allegations in the complaint – for example, the allegation that "Anthropic knowingly trained its AI models on infringing content on a massive

scale in order to enable those models to generate responses to user prompts that infringe Publishers' copyrighted lyrics." Dkt. 1, ¶ 122; *see* Dkt. 328 at 1, 4 (citing Dkt. 1, ¶¶ 8, 108, 122). Anthropic has argued that it needs access to unrelied-upon prompts and outputs in order to rebut such allegations. Dkt. 328 at 4 ("This argument fails if those successes required hundreds of unsuccessful attempts to coax purportedly infringing outputs from Claude. Anthropic thus has a compelling need for the full picture."). But Publishers' prompts and outputs are cited only as examples of *infringement*; at this early stage, it remains unclear what evidence Publishers will rely on to prove the "ease of use" or "massive use" allegations. *See* Dkt. 337, ¶¶ 8, 156, 170 (paragraphs corresponding to ¶¶ 8, 108 and 122 of the original complaint). Thus, at this stage of the litigation, Anthropic's discovery requests and the attendant waiver argument are overbroad. RFP no. 38, for example, requests "Documents sufficient to show each of the prompts [Publishers] entered into Claude … regardless of whether the intent was to generate allegedly infringing, including … the user or organization account used, any system prompts and/or temperature setting sued, and the date and time on which the prompt was entered." This calls for a sweeping subject-matter waiver for all unrelied-upon prompts and outputs and goes too far regardless of whether the prompts, settings and corresponding output are fact work product or opinion work product. Thus, given the broad nature of Anthropic's requests, requiring the production of all undisclosed prompts, settings and outputs is neither "closely tailored" to Anthropic's needs nor limited to what is necessary under the fairness principle. *See Sanmina*, 968 F.3d at 1124.

This Court therefore **DENIES** Anthropic's requests **without prejudice** to Anthropic seeking to discover, as necessary, specific facts supporting specific contentions or opinions disclosed in the course of the litigation.

////

////

////

////

////

II. DKT. NO. 341: SAMPLING PROTOCOL FOR RFP NOS. 50-51

This dispute arises from Publishers' requests for prompts and outputs from Claude products that relate to song lyrics, for which this Court previously ordered that a "statistically significant sample would be proportional to the needs of the case." *See* Dkts. 306, 318. The Parties have submitted competing proposals along with expert declarations, and the Court has had an opportunity to question the experts. *See* Dkts. 341, 341-2, 352, 372.

At the outset, the Court notes that during the hearing, Publishers asked this Court to examine Anthropic's expert, Ms. Chen and strike her declaration because at least one of the citations therein appeared to have been an "AI hallucination": a citation to an article that did not exist and whose purported authors had never worked together. *See* Hrg. Tr. at 7:11-11:25. The Court gave Anthropic time to investigate the circumstances surrounding the challenged citation. Having considered the declaration of Anthropic's counsel and Publishers' response (Dkts. 371, 373), the Court finds this issue is a serious one—if not quite so grave as it at first appeared.

Anthropic's counsel protests that this was "an honest citation mistake" but admits that Claude.ai was used to "properly format" at least three citations and, in doing so, generated a fictitious article name with inaccurate authors (who have never worked together) for the citation at issue. Dkt. 371, ¶¶ 3, 6. That is a plain and simple AI hallucination. Yet the underlying article exists, was properly linked to and was located by a human being using Google search; so, this is not a case where "attorneys and experts [have] abdicate[d] their independent judgment and critical thinking skills in favor of ready-made, AI-generated answers…." *Contra* Dkt. 373 at 2 (quoting Kohls v. Ellison, No. 24-cv-3754 (LMP/DLM), 2025 WL 66514 at *4 (D. Minn. Jan. 10, 2025). A remaining serious concern, however, is Anthropic's attestation that a "manual citation check" was performed but "did not catch th[e] error." Dkt. 372, ¶ 6. It is not clear how such an error—including a complete change in article title—could have escaped correction during manual cite-check by a human being. Furthermore, although the undersigned's standing order does not expressly address the use of AI by parties or counsel, Section VIII.G of Judge Lee's Civil Standing Order requires a certification "that lead trial counsel has personally verified the content's accuracy." J. Lee's Civil Standing Order, § VIII.G (emphasis added). Neither the certification nor

verification has occurred here. In sum, the Court **STRIKES-IN-PART** Ms. Chen's declaration, striking paragraph 9, and notes for the record that this issue undermines the overall credibility of Ms. Chen's written declaration, a factor in the Court's conclusion.

Turning to the substance of the dispute, based upon the Parties' agreed upon parameters, the Court adopts the values of a 95% confidence level corresponding to a "Z-score" of 1.96 and an expected prevalence of 0.00006. Dkt. 352, ¶ 12; Dkt. 341-2, ¶¶ 5, 12. The disagreement between the Parties is with respect to the margin of error: Anthropic offers a 25% margin of error, while Publishers would require a 5% margin. Dkt. 3, 7. During the hearing, the Parties' experts testified that there was room within the 5 to 25% (Publisher's expert Mr. Buchan) or 10 to 50% (Ms. Chen) range—*i.e.*, agreeing that anywhere from 10 to 25% would be statistically valid. *See* Hrg. Tr. at 26:19-28:1, 32:21-33:2. Based upon the foregoing, the Court determines that a margin of error of approximately 11.3% is within the range that will yield a representative sample and will appropriately balance the associated burden against the needs of the case. Calculating a sample from these values according to the experts' formulas, the Court **ORDERS** Anthropic to produce a sample in accordance with the following protocol:

- Anthropic shall produce a total of 5 million prompt-output pairs (*i.e.*, each containing a given prompt and the corresponding Claude output);
- These prompt-output pairs shall be drawn equally from pre-suit and post-suit data, *i.e.*, with 2.5 million records from between September 22 and October 18, 2023, and 2.5 million records from between October 19, 2023 and March 22, 2024; and
- These prompt-output pairs shall be randomly selected from the respective periods.

Anthropic will produce this sample as soon as practicable and **no later than the deadline for substantial completion of remaining document productions**, currently July 14, 2025.

////

////

////

////

////

III. DKT. NO. 345: ANTHROPIC'S CONFIDENTIALITY DESIGNATIONS

With regard to the final dispute, Publishers challenge three categories of "Highly Confidential—Attorneys' Eyes Only" ("HC-AEO") designated materials: (1) Anthropic's 9,418 Claude prompt-output records produced to-date; (2) various Claude use statistics and figures; and (3) two training datasets used by Anthropic. Dkt. 345 at 1. Publishers' concerns regarding over-designation by Anthropic are not unfounded. The protective order ("PO") in this case requires the Parties to "take care to limit any [confidentiality] designations to specific material that qualifies under the appropriate standard [and, to] the extent practicable … designate for protection only those parts of material documents, [etc.] that qualify." Dkt. 293, § V.1. While some level of confidentiality is appropriate for the materials, a blanket HC-AEO designation is not.

First, with regard to Claude prompts and outputs generally, the Court is cognizant not only of the roughly 9,000 prompts produced to-date but also of the forthcoming production of the 5 million record sample. This Court has previously recognized that Anthropic's "users have a privacy interest at stake" in their use of Claude. Dkt. 318. 4. Anthropic does not object to down-designating prompts and outputs to "Confidential" rather than "HC-AEO," and Publishers are unable to identify prejudice apart from the assertion that it makes quoting from prompts and outputs in public filings more difficult. Accordingly, recognizing that it would be impractical if not impossible to require Anthropic to review and redact each prompt-output pair to maintain users' privacy, Anthropic may designate prompts and outputs categorically as "Confidential" at the outset under the PO. However, Anthropic is cautioned that the standard for designation under the PO and this Court's recognition of a general privacy interest are separate from any inquiry into whether documents in the public record should be sealed.[1]

For the Second category, Anthropic's use of the "HC-AEO" designation for Claude use

[1] Under the PO, confidential information is any information that is "nonpublic, sensitive, or proprietary information, the disclosure of which could harm the privacy or business interests of any person." Dkt. 293, § II.6. But a Party seeking to maintain documents under seal must meet either the "good cause" or "compelling reasons" standard. *Ctr. For Auto Safety v. Chrysler Grp.*, 809 F.3d 1092, 1099 (9th Cir. 2016). Accordingly, if Anthropic submits *specific* prompts or outputs in connection with a filing, the burden falls on Anthropic to evaluate whether it can meet the applicable standard for that filing and, if it cannot, de-designate and/or propose tailored redactions for that specific prompt/output record.

statistics and figures has been overbroad. Anthropic did not meaningfully defend its designation of information related to the "sampling numbers" at the hearing, agreeing that such statistics could be down-designated to "Confidential" and subcategorized as needed, but instead pointed to more detailed information in its response to Interrogatory no. 4 that it asserts was properly designated. Hrg. Tr. at 68:23-71:3. Having reviewed Anthropic's submission under seal, (Dkt. 368), the Court agrees that Anthropic's query counts, broken down to such detail, are the type of sensitive metrics that may properly be maintained as "HC-AEO" material. Anthropic shall down-designate its other statistics, including less granular and less comprehensive use-counts, to "Confidential."

Third and finally, the Court agrees with Anthropic that, as a general matter, the training datasets used to train Claude are competitively sensitive and may properly be maintained as "HC-AEO" material. Of course, if Anthropic has publicly disclosed the training datasets at issue, then confidentiality is not appropriate. But Anthropic maintains that it "has never publicly disclosed using either dataset to train Claude." Dkt. 345 at 10. Neither Publishers' citations to what Anthropic's competitors have disclosed about their products, nor their citations to a 2021 article that predates Claude and does not mention Claude, refute this representation. Accordingly, Anthropic may maintain this information as "HC-AEO" material.

SO ORDERED.

Dated: May 23, 2025

SUSAN VAN KEULEN
United States Magistrate Judge

United States District Court
Northern District of California