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 RESOLVING DISCOVERY DISPUTES

Re: Dkt. Nos. 302, 310

Before the Court are two related discovery disputes. The first is set forth in a joint statement regarding Plaintiffs' ("Publishers") Requests for Production ("RFPs") Nos. 50 and 51 directed at specific prompts and outputs from Claude products that relate to song lyrics. Dkts. 302. The Court issued Tentative Rulings regarding this dispute and set the matter for hearing. Dkt. 306 ("Tentative Rulings"). The second dispute also arises from the same RFPs, Nos. 50 and 51, pursuant to which Publishers seek to compel Defendant ("Anthropic") to reveal usernames and email addresses for the prompts and Claude output that it has produced. *See* Dkt. 310 at 2.[1] Both matters came on for hearing on March 18, 2025. Having considered the Joint Statements, the Parties' respective arguments and relevant statutory and case law, the Court resolves the disputes as follows.

I. PUBLISHERS' REQUESTS FOR SEARCHES OF THE CLAUDE DATASET OUTSIDE OF THE 500 WORKS-IN-SUIT

As the Parties explained at the hearing, this case includes a dataset of 6 months of Claude

[1] For both disputes, Plaintiffs have moved for consideration of whether another party's materials should be sealed, as required by Civil L.R. 79-5(f). *See* Dkt. 301 at 2 (explaining that "Publishers do not take any position on whether the information so designated is confidential" and have filed portions of the joint statement under seal at Defendant's request); Dkt. 309 (same with regard to Dkt. 310). As explained in its accompanying order, these motions are GRANTED, but the Court will refer to public information in this Order in the interest of public access.

user prompts and outputs, encompassing a very large number of records[2] from September 2023 to March 2024. Dkt. 302 at 7. The Parties now agree that "evidence regarding Claude's typical uses … could be relevant," even though searches for such evidence reach beyond the 500 Works-In-Suit. Dkt. 302 at 9; *see also* Dkt. 306 at 1-2. However, searches over the entire dataset present issues of technical infeasibility and burden. Dkt. 302 at 6-7. In light of these challenges, the Court indicated to the Parties in its Tentative Rulings that a statistically significant sample would be proportional to the needs of the case. Dkt. 306 at 2. As directed, the Parties met and conferred to develop a sampling protocol and submitted competing proposals which were discussed at length in the hearing. Although both proposals have some merit, the Court is persuaded that Anthropic's approach, with the following guidelines, will allow the Parties to harvest the necessary information from the dataset in an efficient, effective and timely manner. Fed. R. Civ. P. 1.

Accordingly, the Court **ORDERS** Anthropic to produce a statistically significant sample of the Claude prompt and output records. The Parties must meet and confer promptly to determine the size and components of the sample. At a minimum, however, the sample must include both pre-suit and post-suit prompts and outputs and must not separate outputs from their prompts. Anthropic shall be precluded from asserting that the sample is not a statistically significant representation of the broader 6-month dataset. If the Parties are unable to reach agreement, they are to succinctly delineate the specific points of dispute in a Joint Discovery Submission in accordance with this Court's standing order **no later than April 4, 2025**. In addition, each side may submit a single expert declaration, whether from an in-house or independent expert, in support of their respective positions. Any such declaration must closely track the delineation of issues set forth in the Joint Submission and shall not exceed 10 pages. The Parties may stipulate to a modest extension of this deadline if there is a good faith belief that they may reach agreement.

///

///

[2] The precise number of records is under seal. Dkt. 301-2 at 6.

II. PUBLISHERS' REQUESTS FOR SEARCHES AS TO THE 500 WORKS-IN-SUIT

Next, the Parties dispute what searches are reasonable to run against the Claude dataset with regard to the 500 Works-In-Suit. Dkt. 302 at 1-2, 5-7. At the hearing the Parties represented that they had reached agreement as to searches to be run for the 183 Works-In-Suit where the names of the works included common or "noise" words.[3] The Parties remained at an impasse as to how to construct searches for the 317 Works-In-Suit whose titles were unique enough be included as search terms themselves.[4] Publishers argued that the proper search to be run was "lyrics" AND "[the name of the Work-In-Suit]." Anthropic argued that such a search was overly broad and that a "within 5" proximity connector between "lyrics" and "[the name of the Work-In-Suit]" was more appropriate.

The Court agrees that a simple "and" connector is likely to be overbroad in this case but is also concerned that, as Publishers explained, there are relevant prompts and outputs wherein the name of the Work-In-Suit is greater than 5 words away from "lyrics." Accordingly, the Court adopts Publishers' compromise proposed at the hearing and **ORDERS** that Anthropic shall use a "within 20" proximity connector for its Works-In-Suit searches.

III. PUBLISHERS' REQUEST FOR USERNAME AND EMAIL INFORMATION

Finally, in the second joint statement, which was not addressed in the Court's Tentative Rulings but was taken up at the hearing, Publishers seek usernames and email addresses of users who prompted and received output from Claude. Dkt. 310 at 2. In support, Publishers make essentially four arguments: first, that Anthropic may argue that it is the individual users, not Anthropic itself, who are liable for copyright infringement and thus Publishers are entitled to know who those users are (*id.* at 3); second that some Claude users "may possess other critical information to Publishers' claims," such as earlier prompts and outputs in their Claude accounts

---

[3] Such titles include, for example, "With You." Dkt. 302 at 10. Anthropic also explains that "'noise' words like 'me,' 'you,' 'a,' 'the,' and 'to'" are all "excluded by default" from its searchable indices. Dkt. 302 at 8.

[4] The Parties did not identify a specific example of such a work in the Joint Statement, but "Uptown Funk," which is one of the Works-In-Suit, (*see* Dkt. 1-3 at line 26), appears to be a likely exemplar of the 317 works.

(*id.*); third, that Publishers must be able to "differentiate between prompts by Anthropic's employees and agents, … third-party Claude users, … and Publishers [themselves]" (*id.* at 3-4); and fourth, to understand damages-related information such as subscriptions, willfulness, etc., (*id.* at 4). Anthropic does not argue that user information is irrelevant but contends that it cannot disclose such information under the Stored Communications Act ("SCA") and that disclosure is disproportionate to the needs of the case when considering the users' competing privacy interests. *Id.* at 7-10. Publishers disagree that the SCA applies and contend that Anthropic's users do not expect privacy in their interactions with Claude. *Id.* at 4-6.

The Court need not decide whether or to what extent the SCA applies at this juncture because the Court finds that the users' privacy interests, when balanced against Publishers' need for relevant information, support some redactions. As an initial matter, the Court agrees with Anthropic that its users have a privacy interest at stake and that such "privacy interests are not extinguished simply because Anthropic advises it 'may' disclose certain personal data to third parties." *Id.* at 10. Nor, however, are these privacy interests enough to deny discovery into usernames and emails entirely. *See Satmodo, LLC v. Whenever Commc'ns, LLC*, 2018 WL 3495832, at *6 (S.D. Cal. July 20, 2018) (while the court refused to order production of unredacted printouts and browser histories as "intrusive" and "disproportionate to the needs of the case," it explained that this was in part because they were "unlikely to yield any information due to the length of time browser histories are generally saved."). While the privacy interests here are commensurate with those in *Satmodo*, some of Publishers' needs are more compelling. For example, the Court agrees that Publishers must be able to differentiate between themselves, Anthropic's own employees and agents and third parties; and, in limited circumstances, Publishers may be able to show a need for information as to *specific* users, based on the prompts and outputs related to those users. Accordingly, the Court **GRANTS-IN-PART** and **DENIES-IN-PART** Publishers' request for usernames and emails and **ORDERS** as follows:

- Anthropic shall make available the usernames and emails for any of its own employees and agents;

///

- Anthropic shall make available the usernames and emails for Publishers' prompts and outputs of which it is aware;
- Anthropic may anonymize third-party usernames and emails while linking the anonymized usernames and emails to their prompts and outputs; and
- Anthropic must maintain for the duration of this litigation a "key" or some other method to enable it de-anonymize usernames or emails.

SO ORDERED.

Dated: March 25, 2025

SUSAN VAN KEULEN
United States Magistrate Judge

United States District Court
Northern District of California