

OPPENHEIM + ZEBRAK, LLP

Matthew J. Oppenheim
Nicholas C. Hailey
Audrey L. Adu-Appiah
(admitted *pro hac vice*)
4530 Wisconsin Ave., NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
matt@oandzlaw.com
nick@oandzlaw.com
aadu-appiah@oandzlaw.com

Jennifer Pariser
Andrew Guerra
Timothy Chung
(admitted *pro hac vice*)
461 5th Avenue, 19th Floor
New York, NY 10017
Telephone: (212) 951-1156
jpariser@oandzlaw.com
andrew@oandzlaw.com
tchung@oandzlaw.com

Attorneys for Plaintiffs

COBLENTZ PATCH DUFFY & BASS LLP

Jeffrey G. Knowles (SBN 129754)
One Montgomery Street, Suite 3000
San Francisco, CA 94104
Telephone: (415) 391-4800
ef-jgk@cpdb.com

COWAN, LIEBOWITZ & LATMAN, P.C.

Richard S. Mandel
Jonathan Z. King
Richard Dannay
(admitted *pro hac vice*)
114 West 47th Street
New York, NY 10036-1525
Telephone: (212) 790-9200
rsm@ccl.com
jzk@ccl.com
rxd@ccl.com

**UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION**

CONCORD MUSIC GROUP, INC., ET AL.,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

Case Number: 5:24-cv-03811-EKL-SVK

**PLAINTIFFS' OPPOSITION TO
ANTHROPIC'S DECLARATION IN
SUPPORT OF SEALING (ECF NO. 343)**

REDACTED

Judge Eumi K. Lee
Magistrate Judge Susan van Keulen

Pursuant to L.R. 79-5(f)(4), Plaintiffs (“Publishers”) respectfully submit this Response in Opposition to Anthropic’s Declaration in Support of Publishers’ Administration Motion to Consider Whether Another Party’s Material Should Be Sealed, ECF No. 343.

Anthropic fails to identify any legitimate basis for sealing Publishers’ First Amended Complaint (“FAC”), ECF No. 337. Anthropic seeks to seal (1) a short excerpt from a single Claude user prompt requesting lyrics to Publishers’ song in the most general terms, and quotations of two Claude outputs comprised almost entirely of Publishers’ own copyrighted lyrics, none of which include any sensitive user information; and (2) the general fact that Anthropic used two common AI training datasets, information that is already public and that would not cause Anthropic competitive harm. Nothing in the limited language at issue implicates any legitimate privacy interests or threatens any competitive harm to Anthropic sufficient to warrant sealing.

I. Anthropic does not articulate any compelling reasons for sealing the FAC.

The Ninth Circuit recognizes “a strong presumption in favor of access to court records.” *Foltz v. State Farm Mut. Auto. Ins. Co.*, 331 F.3d 1122, 1135 (9th Cir. 2003). “It is well-established that the fruits of pretrial discovery are, in the absence of a court order to the contrary, presumptively public.” *San Jose Mercury News, Inc. v. U.S. Dist. Ct.—N. Dist. (San Jose)*, 187 F.3d 1096, 1103 (9th Cir. 1999). This Court has recognized that “public access to court records is a fundamental aspect of our democracy and justice system.” Standing Order at 5.

As the party seeking to seal Publishers’ FAC, Anthropic “bears the burden of overcoming this strong presumption [of public access] by meeting the ‘compelling reasons’ standard,” which requires Anthropic to “articulate[] compelling reasons supported by specific factual findings that outweigh the general history of access and public policies favoring disclosure.” *Kamakana v. City & Cnty. of Honolulu*, 447 F.3d 1172, 1178 (9th Cir. 2006) (quoting *Foltz*, 331 F.3d at 1135). As part of that showing, Anthropic must demonstrate “(i) the legitimate private or public interests that warrant sealing; (ii) the injury that will result if sealing is denied; and (iii) why a less restrictive alternative to sealing is not sufficient.” L.R. 79-5(c)(1). Boilerplate allegations of harm and “hypothesis or conjecture” are insufficient. *Ctr. for Auto Safety v. Chrysler Grp., LLC*, 809 F.3d 1092, 1097 (9th Cir. 2016) (quoting *Kamakana*, 447 F.3d at 1179).

1 As detailed below, Anthropic has not met this “stringent standard” with respect to the
 2 general, non-sensitive language it seeks to seal here. *Id.* Its sealing request should be denied.

3 **II. There is no compelling reason to seal the limited, anonymous Claude prompt and**
 4 **output excerpts in the FAC.**

5 A review of the specific Claude prompt and output excerpts at issue makes clear that there
 6 is no compelling reason whatsoever to seal this information. Specifically, Anthropic seeks to seal
 7 portions of a single user prompt that requests the lyrics to Publishers’ work “The Boys Are Back
 8 in Town” in general language, without including any sensitive user information (FAC at 32:26,
 9 33:1); and two Claude outputs comprised almost entirely of Publishers’ own copyrighted lyrics—
 10 to “The Boys Are Back in Town,” “Lean on Me,” and “Life is a Highway”—again without any
 11 sensitive user information (FAC at 33:4-34:9, 39:14-17). Anthropic fails to show any privacy
 12 interest that could possibly require sealing this general language, it fails to demonstrate any
 13 specific injury that would result from disclosure of these anonymous excerpts, and it fails to
 14 explain why a less restrictive alternative to sealing is not available.

15 Indeed, Anthropic never identifies what user details or supposedly private user information
 16 are actually contained in these specific prompt and output excerpts that would warrant sealing.
 17 Contrary to Anthropic’s boilerplate claims that these Claude prompt and output excerpts include
 18 “sensitive, nonpublic” information, nothing in these limited excerpts could possibly reveal any
 19 sensitive user information or “result in an invasion of [a] third party’s privacy.” Likewise, while
 20 Anthropic claims that sealing is necessary because the Claude prompt and output excerpts “could
 21 be used to tie the data to a specific user,” ECF No. 343 ¶ 6, no personally identifiable information
 22 or details could conceivably be discerned from this limited language. Separately, Anthropic fails
 23 to articulate any basis for sealing Claude-generated output that simply copies lyrics *owned by*
 24 *Publishers themselves* and does not include any sensitive user information.

25 Even if certain private information were included in the Claude prompts and output in the
 26 FAC (which it is not), that would not necessarily warrant sealing. That is because Anthropic has
 27 already anonymized every single one of its Claude records—removing user names and email
 28 addresses before producing them to Publishers—such that they cannot be tied to any specific users.

1 What's more, Anthropic has gone even further and—without seeking the Court's
2 permission—unilaterally “masked” references to other individual names in its Claude prompt and
3 output records in a way that cannot be undone. That includes permanently removing *non-user*
4 names such as the names of songwriters whose lyrics Anthropic is infringing and names that appear
5 in Publishers' lyrics themselves (as reflected in Anthropic's alteration of the lyrics to “The Boys
6 Are Back in Town,” FAC at 33:18-20, 33:25-26), making certain records indecipherable.

7 Thus, even more detailed excerpts of Claude prompts and output than the ones at issue here
8 would not necessarily implicate legitimate user privacy interests sufficient to require sealing.

9 In sum, nothing in these limited excerpts could possibly “gratify private spite, promote
10 public scandal, circulate libelous statements, or release trade secrets” such that sealing might be
11 warranted. *Kamakana*, 447 F.3d at 1178 (citing *Nixon v. Warner Commc'ns, Inc.*, 435 U.S. 589,
12 598–99 (1978)). Anthropic's “[b]road allegations of harm, unsubstantiated by specific examples
13 of articulated reasoning” cannot carry its burden of demonstrating a compelling reason to seal this
14 information. *Beckman Indus., Inc. v. Int'l Ins. Co.*, 966 F.2d 470, 476 (9th Cir. 1992). That other
15 Claude prompts and output might include sufficiently private user information to warrant sealing
16 is no reason to seal the limited non-sensitive excerpts at issue here. *See, e.g., Bunsow De Mory*
17 *LLP v. N. Forty Consulting LLC*, 2020 WL 7872199, at *2 (N.D. Cal. Sept. 21, 2020) (rejecting
18 sealing request where “arguments in favor of sealing the third parties' identities relie[d] on the
19 context and content of other potentially sealable material in the complaint rather than the sealable
20 nature of the parties' names, and [did] not offer a compelling reason for sealing the names of these
21 third-parties themselves”).

22 Finally, it bears emphasizing that sealing is also unwarranted because the specific language
23 of these Claude prompt and output excerpts goes directly to Publishers' claims for direct and
24 secondary copyright infringement against Anthropic. To rebut Anthropic's contention that no
25 “typical” users seek lyrics from Claude, Publishers included in its FAC specific allegations and
26 examples of Claude users doing just that. Likewise, to address Anthropic's motion to dismiss
27 arguments and the Court's ruling on that motion, *see e.g.*, ECF No. 322 at 5–8, Publishers included
28 specific allegations and examples of third-party Claude users prompting the model for Publishers'

lyrics and receiving output copying those lyrics. Thus, this language is far from merely being “tangentially related” to the “underlying cause of action.” *Kamakana*, 447 F.3d at 1179. The language matters to Publishers’ claims. That weighs further against sealing. *Id.* at 1179–80.

III. There is no compelling reason to seal the fact that Anthropic used two common AI training datasets.

Nor does Anthropic identify any compelling reasons to seal the general fact that its training corpus for Claude includes [REDACTED] datasets.

As an initial matter, the sentence Anthropic seeks to seal says nothing about Anthropic’s specific “proprietary training process” or its specific “dataset mixes,” as Anthropic claims. ECF No. 343 ¶ 5. Rather, the language simply refers to the fact that Anthropic includes two common AI training datasets in its overall Claude training corpus, without further discussion of any specific training process or dataset mix it employs between those and other datasets. That is no trade secret.

Moreover, disclosure of the fact that Claude’s training corpus includes these datasets cannot possibly “harm [Anthropic’s] competitive standing,” because this information is already public. For instance, Anthropic has publicly disclosed (1) training its AI models on the [REDACTED] dataset,¹ which is known to be an [REDACTED];² and (2) training its commercial Claude models on the Pile dataset,³ of which [REDACTED] is a known sub-dataset.⁴ It is well-established that sealing is not warranted where “the information has already become public,” as it is here. *Rodman v. Safeway, Inc.*, 2013 WL 12320765, at *2 (N.D. Cal. 2013) (denying request to seal materials that have become public); *see also* Sealing Order (Dec. 19, 2024), ECF No. 282 (the Court’s denying certain sealing requests where information was already part of the public record).

Additionally, no competitive harm could arise from disclosure of this general information

¹ See, e.g., AMANDA ASKELL ET AL., A GENERAL LANGUAGE ASSISTANT AS A LABORATORY FOR ALIGNMENT (2021) 27, <https://arxiv.org/pdf/2112.00861>.

² See, e.g., [REDACTED]; [REDACTED]; *see also* Am. Compl. ¶¶ 71-72, ECF No. 337.

³ See, e.g., Am. Compl. ¶¶ 69-70, ECF No. 337.

⁴ See, e.g., Annie Gilbertson and Alex Reisner, *Apple, Nvidia, Anthropic Used Thousands of Swiped YouTube Videos to Train AI*, PROOF NEWS (Jul. 16, 2024), <https://www.proofnews.org/apple-nvidia-anthropic-used-thousands-of-swiped-youtube-videos-to-train-ai/>; LEO GAO ET AL., THE PILE: AN 800GB DATASET OF DIVERSE TEXT FOR LANGUAGE MODELING (2020) 4, <https://arxiv.org/pdf/2101.00027>.

1 because these AI training datasets are already used by other AI companies. For instance, it is public
 2 knowledge that OpenAI's ChatGPT-3, a close competitor of Anthropic's Claude models, was
 3 trained on the [REDACTED] dataset,⁵ [REDACTED].

4 "Information does not have value to a competitor merely because the competitor does not have
 5 access to it." *Travelers Property Cas. Co. v. Centex Homes*, 2013 WL 707918, at *1 (N.D. Cal.
 6 2013). Where Anthropic's competitors already train on these allegedly-secret datasets, "it is
 7 unclear how [those] competitors could derive any economic benefit" from this knowledge. *Id.*

8 Finally, this general information about Anthropic's AI training is a matter of public interest
 9 and should not be sealed. The high-level makeup of Anthropic's training corpus bears directly on
 10 Publishers' core claims, including Publishers' copyright management information removal claims,
 11 e.g., FAC ¶ 72. In another AI copyright infringement case, Judge Chhabria rejected similar efforts
 12 by Meta to conceal general references to third-party datasets on which Meta trained its AI models:

13 It is clear that Meta's sealing request is not designed to protect against the disclosure
 14 of sensitive business information that competitors could use to their advantage.
 Rather, it is designed to avoid negative publicity.

15 *Kadrey v. Meta Platforms, Inc.*, No. 23-cv-03417-VC (N.D. Cal. 2025), ECF No. 373.

16 Unfortunately, Anthropic's overbroad request to seal this non-sensitive information is
 17 emblematic of the approach it has taken to confidentiality designations in this case more broadly.
 18 In fact, Anthropic has designated a full 96% of the 13,189 documents it has produced so far as
 19 "Highly Confidential—Attorneys' Eyes Only." This overbroad approach unduly impedes
 20 Publishers' ability to utilize discovery relevant to its claims, and unnecessarily burdens the Court
 21 and Publishers. *See, e.g., United States v. Safran Grp., S.A.*, 2018 WL 11436322, at *4 (N.D. Cal.
 22 Sept. 14, 2018) (recognizing that motions to seal "inflict heavy burdens on courts").

23 For all of these reasons, Anthropic's request to seal these materials should be denied.
 24
 25
 26
 27
 28

⁵ BROWN ET AL., LANGUAGE MODELS ARE FEW-SHOT LEARNERS (2020) 27, <https://arxiv.org/pdf/2005.14165>.

1 Dated: May 6, 2025

Respectfully submitted,

2 /s/ Timothy Chung

3 **OPPENHEIM + ZEBRAK, LLP**

4 Matthew J. Oppenheim

Nicholas C. Hailey

5 Audrey L. Adu-Appiah

(admitted *pro hac vice*)

6 4530 Wisconsin Ave., NW, 5th Floor

Washington, DC 20016

7 Telephone: (202) 480-2999

matt@oandzlaw.com

8 nick@oandzlaw.com

aadu-appiah@oandzlaw.com

9 Jennifer Pariser

10 Andrew Guerra

11 Timothy Chung

(admitted *pro hac vice*)

12 461 5th Avenue, 19th Floor

New York, NY 10017

13 Telephone: (212) 951-1156

jpariser@oandzlaw.com,

14 andrew@oandzlaw.com

tchung@oandzlaw.com

15 **COBLENTZ PATCH DUFFY & BASS LLP**

16 Jeffrey G. Knowles (SBN 129754)

17 One Montgomery Street, Suite 3000

San Francisco, CA 94104

18 Telephone: (415) 391-4800

ef-jgk@cpdb.com

19 **COWAN, LIEBOWITZ & LATMAN, P.C.**

20 Richard S. Mandel

Jonathan Z. King

21 Richard Dannay

(admitted *pro hac vice*)

22 114 West 47th Street

New York, NY 10036-1525

23 Telephone: (212) 790-9200

24 rsm@cjl.com

jzk@cjl.com

25 rxd@cjl.com

26 *Attorneys for Plaintiffs*