

1 [Counsel on signature page]

2 **UNITED STATES DISTRICT COURT**
3 **NORTHERN DISTRICT OF CALIFORNIA**
4 **SAN JOSE DIVISION**

5 **CONCORD MUSIC GROUP, INC., ET AL.,**

6 **Plaintiffs,**

7 **v.**

8 **ANTHROPIC PBC,**

9 **Defendant.**

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

**JOINT DISCOVERY SUBMISSION
REGARDING DISPUTE AS TO
SAMPLING PROTOCOL TO ADDRESS
PUBLISHERS' RFP NOS. 50-51**

REDACTED

Judge Eumi K. Lee
Magistrate Judge Susan van Keulen

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

1 Pursuant to the Court's March 25, 2025 Order, Dkt. 318, Section 8 of the Court's Civil
2 Standing Order, and L.R. 37-1 and 37-2, Plaintiffs ("Publishers") and Defendant Anthropic PBC
3 ("Anthropic") respectfully submit this Joint Discovery Submission regarding the development of
4 a sampling protocol to resolve the dispute regarding Publishers' RFPs 50-51. The Parties' lead
5 counsel met and conferred via videoconference (including March 31 and April 14, 2025) and email,
6 but the Parties are unable to reach agreement. The current fact discovery deadline is in 134 days.

7 **I. Publishers' Position**

8 Despite Publishers' proposing several different compromises, the Parties have
9 unfortunately been unable to reach agreement on the size and components of a statistically
10 significant sample of Claude prompts and output sufficient to address Publishers' RFPs 50-51.

11 Although Anthropic proposed to the Court at the Mar. 18 discovery hearing that the Parties
12 "pick a big enough sample" so that "both sides can run whatever searches they want," Hr'g Tr.
13 17:20-21, 21:21-23 (Mar. 18, 2025), it has since tried to walk back that commitment considerably,
14 seeking to shrink the sample's size and undermine its utility to Publishers. Anthropic's current
15 proposal is not aimed at allowing Publishers a meaningful opportunity to search and analyze
16 Claude prompts and output (including, in particular, the various lyric-related records requested by
17 RFPs 50-51) or draw statistically significant conclusions from that analysis. Rather, Anthropic's
18 proposal is aimed only at buttressing its own fair use defense. That is not the right approach.

19 Publishers' proposal, by contrast, is in line with Anthropic's prior commitment to produce
20 a sufficiently large sample such that each party can analyze the data for multiple purposes.
21 Specifically, Publishers propose that Anthropic produce a sample dataset of 24.5 million Claude
22 prompt and output records, or the rough equivalent of (1) [REDACTED] days between Sept. 22 and Oct.
23 18, 2023 (pre-lawsuit), and (2) [REDACTED] days between Oct. 19, 2023 and Mar. 22, 2024 (post-lawsuit).
24 Such a dataset represents a modest [REDACTED] % of the total population of [REDACTED] million records Anthropic
25 has preserved for this six-month period. As detailed below, Publishers' approach will ensure the
26 sample is sufficiently large to draw statistically significant conclusions across a range of
27 complicated issues important to both sides—not just Publishers' claims or Anthropic's defenses.

28 Anthropic's much more limited proposed sample—1 million records, a mere [REDACTED] % of the

1 total records—is too small for the Parties to draw sufficiently reliable conclusions across the data.

2 **A. Publishers’ proposed sample size will allow statistically significant analysis**
 3 **and conclusions as to a range of different types of prompts and output.**

4 At the outset, it bears emphasizing that, not only is the Claude dataset very large, but the
 5 data is complicated and the analyses varied, necessitating the larger sample size Publishers propose.
 6 Publishers’ RFPs 50-51 seek Claude prompts and output that generally “relate to song lyrics,” ECF
 7 No. 318 at 1, but within this category are myriad, overlapping sub-categories of relevant prompts
 8 and output, the precise details of which will be unknown until the Parties review the sample.
 9 Likewise, the Parties may analyze various other types of prompts and output as part of the sample.

10 Some types of Claude prompts and output to be analyzed may occur in relatively large
 11 numbers. For example, out of the [REDACTED] million total records at issue, upwards of 6.5 million contain
 12 the term “lyric.” (While Anthropic has repeatedly refused to search for the term “lyric” across the
 13 full dataset, this estimate is based on Anthropic’s representation that, for a nine-day period in Sept.
 14 2023, the term “lyric” appeared in 170,077 of [REDACTED] million records—[REDACTED] % of all records.) That will
 15 include a wide range of prompts and output—such as prompts specifically requesting Publishers’
 16 and others’ song lyrics, output copying lyrics even when the user did not ask, and records in which
 17 the term “lyric” refers to something other than song lyrics. There are also likely millions of
 18 additional records in which users seek or receive lyrics to Publishers’ works and other songs, but
 19 the specific term “lyric” does not appear (e.g., “What are the words to I Will Survive?”, “What is
 20 the first verse to What a Wonderful World?”, “Write me a song about the death of Buddy Holly”).

21 At the other end of the spectrum, there are Claude prompts and output that—while much
 22 less frequent relative to the full dataset—are nevertheless vitally important to the case, such as:

- 23 • Instances in which one of *Anthropic’s own founders uses the AI model to seek lyrics*, e.g.,
 24 Anthropic_0000016458 (Anthropic co-founder Tom Brown querying, “@Claude what are
 the lyrics to desolation row by [Bob] Dylan?”); Am. Compl. ¶ 113 (alleging same); and
- 25 • Instances in which Claude copies Publishers’ lyrics in output and then, when specifically
 26 asked by the user whether those lyrics are copyrighted, Claude *falsely claims* the lyrics are
 “entirely fictional and written by myself” and “free of any copyrighted material,” *id.* ¶ 104.

27 Prompts and output like these—even if one in a million—are critically relevant to Publishers’
 28 claims and Anthropic’s willful infringement. They must be accounted for in deciding a sample size.

1 This variance in the prevalence of prompts and output to be analyzed, the overlapping and
2 subjective nature of the data, and the fact that much about the data to be analyzed remains unknown
3 at this time all favor a larger sample size to ensure statistical significance. Ex. 1, Buchan Decl. ¶¶
4 13-14, 19, 37. Publishers’ proposed 24.5-million-record sample will better allow the Parties to
5 analyze both frequent and infrequent occurrences alike, ensuring any conclusions derived from the
6 sample are more reliable and precise. *Id.* ¶¶ 30-37; *In re Countrywide Fin. Corp. Mortg.-Backed*
7 *Sec. Litig.*, 984 F. Supp. 2d 1021, 1033 (C.D. Cal. 2013) (“[T]o draw reliable conclusions about a
8 population based on a statistical sample, the sample size must be large enough to support those
9 conclusions.”). The larger the sample, the more precise the conclusions. Buchan Decl. ¶¶ 14, 37.

10 By comparison, Anthropic’s proposed 1-million-record sample is too small to account for
11 the more rare occurrences—such as Anthropic’s founder using Claude to search for Publishers’
12 lyrics, the ***very thing Anthropic denied*** designing Claude to do—that are still central to the case.

13 Anthropic also seeks to apply an “assumed prevalence rate” of 0.006% across the board.
14 But that one-size-fits-all approach is inappropriate here. As explained above, assuming any single
15 prevalence rate vastly oversimplifies the many unique types of prompts and output and different
16 issues to be analyzed. Publishers’ larger sample size takes these considerations into account.

17 Publishers’ proposed sample size is also more consistent with commonly accepted
18 surveying and sampling practices. Publishers’ sample size calculation is based on the following:

- 19 • 95% confidence level: Publishers understand that the Parties are in agreement on this figure.
- 20 • 5% relative margin of error: This margin of error is commonly utilized for sampling
21 purposes, including rare event sampling, and comes with the added benefit of increasing
22 the precision of any estimates derived from the sample data. Buchan Decl. ¶ 29.
- 23 • Anthropic’s “assumed prevalence rate” of 0.006%: Publishers are willing to accept,
arguendo, Anthropic’s assumed prevalence rate for purposes of calculating a sample size.

24 Based on these figures, a statistically significant sample is 24.5 million records. *Id.* ¶¶ 34, 36.

25 By contrast, Anthropic’s proposed 1-million-record sample utilizes the same 95%
26 confidence level and 0.006% prevalence rate, but is premised on a much larger and less precise
27 25% relative margin of error. A 25% margin of error is not appropriate here and is on the very
28 outer threshold of standard practice. *Id.* ¶¶ 30-32. This inflated margin of error appears designed

1 to reduce the sample's size at the cost of undermining the sample's utility. *See id.* ¶ 37.

2 **B. Publishers' proposed approach will also allow statistically significant**
3 **analyses of Claude prompts and output both pre-suit and post-suit.**

4 It is also imperative that the sample allow the Parties to draw reliable conclusions about
5 Claude prompts and output from both (1) before Publishers filed suit, when Anthropic's limited
6 guardrails that were largely ineffective in preventing infringing output, and (2) after Publishers
7 filed suit, when Anthropic adopted additional post-litigation guardrails in response to Publishers'
8 claims. Accordingly, Publishers propose that half of the sample be drawn from Anthropic's [REDACTED]
9 million Claude records from Sept. 22 to Oct. 18, 2023 (pre-lawsuit), and half from its [REDACTED] million
10 records from Oct. 19, 2023 to Mar. 22, 2024 (post-lawsuit). This approach is consistent with the
11 Court's guidance that "the sample must include both pre-suit and post-suit prompts and outputs."
12 ECF No. 318 at 2. By contrast, Anthropic's position that prompts be drawn randomly from across
13 the six-month period for which it preserved and agreed to produce records will skew the sample
14 to the post-suit period, when Claude prompts and output may be very different than during the pre-
15 suit period, in a number of ways. That Anthropic's retention policy did not provide for preserving
16 records for a longer period pre-suit should not allow it to distort the sample in this manner.

17 **C. Publishers' proposed sampling approach will not unduly burden Anthropic.**

18 In addition to ensuring statistical significance, Publishers' proposal also reasonably
19 balances Publishers' need for this discovery against the burden to Anthropic. While Anthropic
20 may face a slightly greater burden in collecting and producing a larger dataset, it has "not provided
21 any information that indicates that the production of such [records] would be unduly burdensome."
22 *Fed. Trade Comm'n v. Tate's Auto Ctr. of Winslow Inc.*, 2019 WL 1130006, at *4 (D. Ariz. Mar.
23 12, 2019) (directing parties to identify an appropriate statistical sample size). That is particularly
24 true where Anthropic will simply be required to process and produce the data at issue—which is
25 minimally burdensome—rather than undertake any substantive review prior to production.
26 Moreover, Anthropic has acknowledged its systems are capable of searching and processing at
27 least [REDACTED] million Claude records at a time (the equivalent of nine days). Hr'g Tr. 19:7–13 (Mar. 18,
28 2025). Publishers' proposal would require only that Anthropic process 24.5 million records (the

rough equivalent of [REDACTED] days, which can be broken up into smaller portions of nine days or less).

D. Anthropic’s proposed approach will serve only its own ends.

Finally, it bears emphasizing that the reason the Parties are undertaking this sample in the first place is to address Publishers’ RFPs 50-51, which seek production of Claude prompts and output relating to song lyrics. ECF No. 318 at 1. Over the course of the Parties’ discussions regarding this sample, however, it has become increasingly apparent that Anthropic is seeking to exploit this sample as an opportunity to buttress its own fair use defense—by focusing on “overall Claude usage” beyond that at issue in Publishers’ RFPs 50-51—while at the same time limiting production of the very lyrics-related records sought by Publishers’ RFPs. This is improper. Anthropic cannot substitute its own objectives for Publishers’ in this manner. *See, e.g., Friedman v. 24 Hour Fitness USA, Inc.*, No. CV0606282AHMCTX, 2009 WL 10672797, at *1 (C.D. Cal. Jan. 12, 2009) (“It would be unfair to limit Plaintiffs to the 200–member sample provided by the Discovery Order and allow Defendant to use evidence regarding members outside the sample.”).

Given that Publishers’ RFPs 50-51 are the basis for this discovery dispute and the Court’s order that the Parties develop a sampling protocol, and given that Publishers’ RFPs 50-51 specifically seek Claude prompts and output relating to song lyrics, the sample that the Parties undertake must at a minimum allow for Publishers to identify and analyze a sufficient number of the specific lyric-related prompts and output they request. Each such prompt and output is a potentially new and separate act of infringement by Anthropic. Each is relevant to Publishers’ claims of direct infringement, secondary copyright infringement, and removal of copyright management information, as well as Anthropic’s willfulness, damages, and other core issues.

For this purpose, Publishers seek a reasonably-sized sample—24.5 million records—that balances Publishers’ entitlement to the documents they seek against the burden of that discovery.

Critically, Publishers’ proposed sample is sufficiently large to address *both* Publishers’ discovery and Anthropic’s goal of analyzing the sample for its own purposes, erring on the side of caution and with the objective of avoiding further discovery disputes as to this sample. Anthropic’s too-small sample, on the other hand, is geared solely toward propping up its fair use defense.

1 In sum, only Publishers’ proposal will fairly allow both Parties to “run whatever searches
2 they want,” as Anthropic initially committed, and “harvest the necessary information from the
3 dataset in an efficient, effective, and timely manner,” as the Court ordered, ECF No. 318 at 2.

4 **II. Anthropic’s Position**

5 Sampling reduces the burden and expense of discovery regarding voluminous data. *See*
6 *Manual for Complex Litigation* § 11.493 (4th ed. 2004) (“Acceptable sampling techniques, in lieu
7 of discovery and presentation of voluminous data from the entire population, can save substantial
8 time and expense, and in some cases provide the only practicable means to collect and present
9 relevant data.”); *accord Tyson Foods, Inc. v. Bouaphakeo*, 577 U.S. 442, 454–55 (2016)
10 (approving use of “a representative or statistical sample” as a “means to establish or defend against
11 liability”). Plaintiffs sought broad discovery of all Claude prompts and outputs related to song
12 lyrics within a dataset of approximately [REDACTED] records. This Court recognized the “technical
13 infeasibility and burden” of that request and adopted Anthropic’s position that a random sample
14 would “allow the Parties to harvest the necessary information from the dataset in an efficient,
15 effective and timely manner.” Dkt. 318 at 2; *see id.* (ordering Anthropic to “produce a statistically
16 significant sample of the Claude prompt and output records” after meeting and conferring about
17 the sample’s “the size and components”). Anthropic’s proposed sample of 1 million records is
18 more than sufficient to evaluate actual Claude usage patterns. Plaintiffs’ proposal—to simply
19 select [REDACTED] full days of data (over 10 million records)—defies basic statistical principles and flatly
20 contradicts their theory of the case. If users were truly exploiting Claude to reproduce lyrics as
21 often as Plaintiffs claim, Plaintiffs would need only a fraction of this data to prove it.

22 When the parties met and conferred on March 31, 2025, Plaintiffs proposed the use of a
23 “pre-sample sample” to develop a rough estimated prevalence rate for lyric-related prompts and
24 outputs. Under this approach, the parties would have used the estimated prevalence rate from a
25 pilot study of approximately 200,000-260,000 records to calculate, using standard statistical
26 formulas, an appropriate sample size for the entire population. Statistical principles dictate that a
27 small estimated prevalence rate from the pilot study merits a larger sample size, and vice versa.
28 The parties negotiated the details of this pre-sample sample in good faith between March 31 and

1 April 9, at which point Plaintiffs abruptly abandoned this approach and instead proposed, for the
2 first time and without any mathematical rationale, a sample of [REDACTED] full days of data. This proposal
3 would have resulted in a sample size of over 20 million records. Since then, Plaintiffs have
4 modified their proposal to [REDACTED] full days of data—[REDACTED] days pre-suit and [REDACTED] days post—for a sample size
5 still totaling over 10 million records. This proposal is neither scientific nor reasonable.

6 Anthropic has proposed a random sample of 1 million records. Under Anthropic’s proposal,
7 1 million chat logs (i.e., user interactions containing both prompts and outputs) would be selected
8 randomly from across the 6-month universe of data. This sample is large enough to allow
9 meaningful conclusions about the entire dataset, but small enough to significantly reduce the
10 burden and expense of discovery into song-related prompts and outputs. Plaintiffs’ proposal of a
11 fixed sample of 10+ million records selected from [REDACTED] days of data is flawed in two respects: (1) it
12 is far larger than necessary to derive meaningful insights about the population in question, either
13 because it implicitly assumes an incomprehensibly low prevalence rate or uses a needlessly precise
14 margin of error; and (2) its focus on just [REDACTED] days of data—[REDACTED] from before the complaint was filed
15 and [REDACTED] after—injects selection bias and is thus unlikely to be representative.

16 **A. Anthropic’s Sample Size of 1 Million Is An Appropriate Size to Draw Conclusions
Regarding the Overall Population**

17 In order to realize the efficiencies of statistical sampling, a sample should be no larger than
18 necessary to achieve statistical validity. *See* Declaration of Qinnan (Olivia) Chen (“Chen Decl.”),
19 Ex. B ¶ 18. Sample size is typically calculated based on an estimated prevalence in the population
20 of the event being studied and the relative margin of error. *Id.* ¶ 5. If an expected prevalence rate
21 is unavailable, either a pilot study can be conducted to determine one, or a “guesstimate” can be
22 used.¹ Here, the parties’ earlier search term negotiations and prompt and output productions thus
23 far demonstrate the relative rarity of the event being studied (prompts for or outputs containing
24 song lyrics), *see* Dkt. 302 at 6–7, 9, but not the precise prevalence rate.

25 In the absence of a pilot study, Anthropic’s 1 million record sample is sufficiently
26 representative to reflect even extremely rare events. *See* Chen Decl. ¶¶ 6–8. Using a 25% relative

27
28 ¹ *See* Chittaranjan Andrade, *Sample Size and Its Importance in Research*, 42 *Indian J. Psych. Med.* 102,
103 (2020), available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC6970301/>.

1 margin of error—widely accepted for extremely rare events—a sample size of 1 million would
2 validly reflect events in the overall population with prevalence rates as low as 0.006%. Using a
3 more stringent 20% relative margin of error, a 1 million record sample size would still capture
4 events with prevalence rates as low as 0.01%. *Id.* ¶¶ 9–12. Any statistical power gained from a
5 larger sample would be outweighed by the diminished efficiencies involved in running and
6 reviewing the results from a larger sample. *Id.* ¶ 17–18; *see also* Fed. R. Civ. P. 26(b)(1) (courts
7 must consider if “the burden or expense of the proposed discovery outweighs its likely benefit”).

8 To the extent Plaintiffs argue that a sample of 1 million is inadequate because it represents
9 too small a fraction of the overall population, that argument is unmoored from rigorous and
10 accepted statistical methods. *See* Chen Decl. ¶ 5 (setting forth the formula that “represents the
11 fundamental statistical approach to determining the minimum sample size needed to make valid
12 inferences about a very large dataset,” including “the one at issue here”). It is also unsupported by
13 case law in this Circuit, which recognizes that the fraction of a representative sample should
14 decrease as the total population (and burden) increases. *See Guzman v. Chipotle Mexican Grill,*
15 *Inc.*, 2018 WL 6092730, at *3 (N.D. Cal. Nov. 21, 2018) (finding a 5% sample appropriate for a
16 “large” population of 43,000 and noting that courts imposed higher percentage samples for “much
17 smaller” populations); *Heredia v. Sunrise Senior Living LLC*, 2019 WL 7865176, at *6 (C.D. Cal.
18 Oct. 31, 2019) (finding a 15% sample of a 13,000 population “appropriately balances Plaintiffs’
19 discovery needs, Defendant’s burdens, and the privacy rights” of the individuals sampled and
20 noting that a larger sample “would be excessive”); *Nia v. Bank of Am., N.A.*, 2023 WL 2583386,
21 at * 5 (S.D. Cal. Mar. 20, 2023) (considering the significant costs required to “process, review,
22 redact, and produce the proposed sample” in deciding to cut the sample size from 20% to 10%).

23 Plaintiffs’ insistence on a sample *ten times* the size that Anthropic proposes cannot be
24 supported by accepted statistical methods. *See Deutsche Bank Nat’l Tr. Co. v. Morgan Stanley*
25 *Mortg. Cap. Holdings LLC*, 289 F. Supp. 3d 484, 496 (S.D.N.Y. 2018) (“Properly done, statistical
26 sampling is not guesswork”). A sample of this size either assumes (a) an extraordinarily low
27 prevalence rate, far lower than the 0.01% or 0.006% already contemplated by Anthropic’s sample;
28 or (b) an unnecessarily stringent relative margin of error. *Id.* ¶ 17. With respect to (a), Plaintiffs’

1 position throughout this litigation has been the opposite: that Claude was designed to—and
 2 frequently does—respond to prompts requesting song lyrics. *See, e.g.*, Dkt. 1 ¶ 11 (“Anthropic
 3 unlawfully enables, encourages, and profits from massive copyright infringement by its users”);
 4 Dkt. 179 at 2 (“Anthropic intended and expected its AI models to respond to requests for Publishers’
 5 lyrics—as a feature, not a bug”); *and* Dkt. 225 at 10 (pointing to search results for the terms “lyric”
 6 and “song” to posit a “potentially huge volume of requests for lyrics by [Claude] users”). In their
 7 amended complaint filed last week, Plaintiffs allege that “literally millions of Claude prompt and
 8 output records contain the term ‘lyric,’” and that this means that “[c]ountless users . . . have
 9 prompted Claude for lyrics.” Dkt. 337 ¶¶ 9–10 (emphasis added). Plaintiffs now conveniently treat
 10 lyric requests as extraordinarily rare events comparable to rare genetic diseases or lightning strikes.
 11 If their claims have merit (which Anthropic disputes), Anthropic’s sampling proposal is already
 12 more than adequate. And with respect to (b), Plaintiffs cannot explain why the Court must exceed
 13 the margin of error standards that are routinely accepted by statisticians when analyzing rare
 14 events. Chen Decl. ¶¶ 9, 17. Plaintiffs’ proposed sample size should be rejected.

15 **B. Anthropic’s Proposed Sample Is Representative of the Entire Population**

16 To extrapolate the results of a sample to the larger population, the sample “must be
 17 randomly selected.” *United States v. Nunez*, 2021 WL 5494588, at *7 (D. Conn. Nov. 23, 2021).
 18 “Random” means that every datapoint in the population has the same probability of becoming part
 19 of the sample. *See* D. Freedman & D. Kaye, FJC, *Manual on Scientific Evidence* 230 (2011).
 20 Scientific randomness “provides assurance of unbiased estimates.” *Id.*; *see also* Chen Decl. ¶ 19.

21 Anthropic has proposed a true random sample: each Claude user interaction in the full
 22 dataset would have an equal chance of being included. This would ensure the sample includes both
 23 pre- and post-suit chat logs but does not suffer from any biases resulting from hand selecting chat
 24 logs from non-representative periods, e.g., (1) weekends versus weekdays, or vice versa; (2)
 25 months when Claude had a smaller userbase; and (3) periods of anomalous user activity (i.e.,
 26 around the filing of Plaintiffs’ complaint when their agents were “investigating”). Chen Decl. ¶¶
 27 19–22. Anthropic’s proposed sample would be unbiased and representative of the full population.

28 Plaintiffs’ proposal is not random in the rigorous, scientific sense because it suffers from

1 each of the above temporal biases. *Id.* “Hand-picking is almost certain to introduce a substantial
2 amount of selection bias into the sample,” *Rosenbohm v. Celco P’ship*, 2019 WL 2141901, at *2
3 (S.D. Ohio May 16, 2019), *objections overruled*, 2019 WL 13507817 (S.D. Ohio July 24, 2019),
4 and Plaintiffs can offer no rationale, grounded in established statistical methods, for introducing
5 such biases, *see also* Freedman & Kaye, 230 (“Looser definitions of randomness are inadequate
6 for statistical purposes.”). Their proposal is unlikely to produce representative results that would
7 allow extrapolation to the larger population and should be rejected.

8 **C. Regardless of the Contours of the Sample, Privacy Safeguards Are Needed**

9 The parties also disagree on the privacy protections required in connection with this
10 sample. These chat logs represent the private communications of third-party Anthropic users. *See*
11 *generally* Dkt. 310 at 7–11. Unlike the prompts and outputs that Anthropic has produced thus far,
12 however, the prompts and outputs in the sample will not be filtered for relevance. Accordingly,
13 the privacy interests of the users associated with these chat logs are exceptionally strong. *Cf.*
14 *Heredia*, 2019 WL 7865176, at *4 (expressing “concern[] about the privacy interests” of the
15 individuals swept into a sample, and balancing those interests in ordering a sampling procedure).

16 To adequately protect these interests, Anthropic proposed—and Plaintiffs materially
17 agreed to—a set of modest procedures, including: (1) that the sample prompts and outputs as well
18 as any aggregate information and/or analysis would be designated Confidential or HC-AEO under
19 the Protective Order, subject to Plaintiffs’ right to challenge those designations; (2) that the sample
20 prompts and outputs would be de-identified, but to the extent any identifying information was not
21 removed from the record prior to disclosure, Anthropic would be given a reasonable opportunity
22 to remove it; and (3) that the sample would be accessible only on a secure platform.

23 Despite Plaintiffs’ initial assent to a set of privacy-protecting procedures, they withdrew
24 that assent when Anthropic would not agree to their’ proposal of █ full days of data. Anthropic
25 respectfully requests that the Court incorporate these protections into a sampling protocol.

1 Dated: April 30, 2025

Respectfully submitted,

2 By: /s/ Timothy Chung

By: /s/ Sarang V. Damle

3 **OPPENHEIM + ZEBRAK, LLP**

LATHAM & WATKINS LLP

4 Matthew J. Oppenheim
5 Nicholas C. Hailey
6 Audrey L. Adu-Appiah
7 (admitted *pro hac vice*)
8 4530 Wisconsin Ave., NW, 5th Floor
9 Washington, DC 20016
10 Telephone: (202) 480-2999
11 matt@oandzlaw.com
12 nick@oandzlaw.com
13 aadu-appiah@oandzlaw.com

Joseph R. Wetzel (SBN 238008)
joe.wetzel@lw.com
Andrew M. Gass (SBN 259694)
andrew.gass@lw.com
Brittany N. Lovejoy (SBN 286813)
britt.lovejoy@lw.com
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: +1.415.391.0600

14 Jennifer Pariser
15 Andrew Guerra
16 Timothy Chung
17 (admitted *pro hac vice*)
18 461 5th Avenue, 19th Floor
19 New York, NY 10017
20 Telephone: (212) 951-1156
21 jpariser@oandzlaw.com
22 andrew@oandzlaw.com
23 tchung@oandzlaw.com

Sarang V. Damle
(admitted *pro hac vice*)
sy.damle@lw.com
Sara Sampoli (SBN 344505)
sara.sampoli@lw.com
555 Eleventh Street NW, Suite 1000
Washington, DC 20004
Telephone: +1.202.637.2200

24 **COBLENTZ PATCH DUFFY & BASS LLP**

Allison L. Stillman
(admitted *pro hac vice*)
alli.stillman@lw.com
1271 Avenue of the Americas
New York, New York 10020
Telephone: +1.212.906.1747

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

Attorneys for Defendant

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

30 Richard S. Mandel
31 Jonathan Z. King
32 Richard Dannay
33 (admitted *pro hac vice*)
34 114 West 47th Street
35 New York, NY 10036-1525
36 Telephone: (212) 790-9200
37 rsm@cll.com
38 jzk@cll.com
39 rxd@cll.com

Attorneys for Plaintiffs

