OPPENHEIM + ZEBRAK, LLP
Matthew J. Oppenheim
Nicholas C. Hailey
Jeffrey M. Gould
Corey Miller
Keith Howell
(admitted *pro hac vice*)
4530 Wisconsin Ave., NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
matt@oandzlaw.com
nick@oandzlaw.com
jeff@oandzlaw.com
corey@oandzlaw.com
khowell@oandzlaw.com

Alexander Kaplan
Jennifer Pariser
Andrew Guerra
Bret Matera
Timothy Chung
Michelle Gomez-Reichman
(admitted *pro hac vice*)
461 5th Avenue, 19th Floor
New York, NY 10017
Telephone: (212) 951-1156
alex@oandzlaw.com
jpariser@oandzlaw.com
andrew@oandzlaw.com
bmatera@oandzlaw.com
tchung@oandzlaw.com
mgomez-reichman@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@cll.com
jzk@cll.com
rxd@cll.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' REPLY IN SUPPORT OF MOTION TO PARTIALLY EXCLUDE THE OPINIONS OF DR. ANDREW BRAY

REDACTED

Judge Eumi K. Lee
Magistrate Judge Susan van Keulen
Hearing Date: October 21, 2026
Time: 10:00 AM
Courtroom 7

Anthropic intends to use Dr. Bray to tell the jury that only 0.047% of Claude prompts sought lyrics and only 0.014% of outputs contained lyrics. ECF No. 750 ("Opp.") at 3. But Anthropic concedes Dr. Bray did not review the full 5 million record Sample.[1] Instead, Dr. Bray derived his figures from word-search analyses he performed, known as "regular expressions." Regular expressions can be used to find ***known*** patterns of words in large datasets. But the literature warns they are unreliable for identifying instances of a concept that can be expressed in myriad ways, like requests for lyrics submitted by individual users. To reliably use regular expressions to find such requests across millions of records, Dr. Bray would need to anticipate the many, varied ways in which these requests could be phrased. But, as Anthropic concedes, Dr. Bray has only a layman's understanding of how requests for lyrics could be worded, and his analysis was indisputably riddled with errors. Dr. Bray's analysis reflects merely how many documents happened to hit on his chosen search terms, which is not relevant to any claim. Aware that his analysis contained "many" mistakes, Dr. Bray chose not to calculate an error rate, further favoring exclusion. Anthropic mischaracterizes the caselaw to conceal a simple truth: no Court has approved an expert's use of a regular expression analysis as unsupported and error-ridden as Dr. Bray's. The Court should grant Publishers' Motion. ECF No. 644 ("Motion" or "Mot.").

ARGUMENT

I. Regular expressions are not a reliable method of searching for a general concept in a large dataset, and Dr. Bray lacks any expertise to design reliable patterns.

Regular expressions, or "regex," operate like the CTRL+F function on one's computer, but with the ability to search for multiple word patterns at once. ECF No. 645-3 ("Dep. Tr.") at 45:13–18 (regular expressions "search for a particular set of patterns of words, spaces, [and] line breaks"); *see also id.* 46:3–5. They are "reliable" only in the sense that they repeatedly identify documents that include a particular pattern of words. The reliability of any regex analysis thus turns on whether the word patterns searched capture the full scope of what is being searched for. Regular expressions thus can be useful to search for known, fixed text like the lyrics to a song. But the literature warns they are unreliable for "identify[ing] instances of a specific concept" like requests

[1] Terms used in this Reply are defined in Publishers' Motion. ECF No. 644.

for lyrics (ECF No. 645-4), which as Dr. Bray acknowledges can be phrased in "[redacted]" ways (Reply ¶ 34).[2] Yet Dr. Bray uses regular expressions in exactly this unreliable way.

Dr. Bray also lacks the expertise to design word patterns that would reliably capture all requests for lyrics. In creating regular expressions, Dr. Bray had to design the patterns to search for lyrics requests entirely from scratch. Anthropic concedes that no one before Dr. Bray has used regular expressions to identify requests for lyrics in any dataset, and that Dr. Bray never performed this task prior to this litigation. As a professor of statistics, Dr. Bray has no expertise regarding how Anthropic's millions of users might request lyrics.[3] Thus, Dr. Bray began his analysis with "just kind of a sense of what [prompts seeking lyric regurgitation] would look like." Dep. Tr. 100:17–24. Such untethered whimsy lacks any reliable basis and strays far from the "reasonable confines of his subject area: statistics and data science." Opp. 7.

Anthropic's main response that the use of regular expressions is "well-established" in computer science, *id.* at 1, misses the point. The problem is not that regular expressions do not reliably return documents that hit on the search terms used, but that there is indisputably no reliable basis for the specific patterns Dr. Bray used to search for prompts seeking and outputs providing lyrics. Anthropic mistakenly focuses on the tool rather than the inputs to the tool. Anthropic's argument is equivalent to arguing that a damages expert's opinion is reliable because she used a calculator, and use of calculators is "well-established" in the financial industry.

The undisputed errors in Dr. Bray's analysis confirm its unreliability. Mot. 6–8. Publishers' Motion identifies dozens of misclassifications by Dr. Bray, none of which Anthropic disputes. *Id.* An analysis that classifies "[redacted]" as a request for lyrics and "[redacted]" as ***not*** a request for lyrics, tells the jury nothing except how many documents happened to hit on Dr. Bray's search terms, a fact irrelevant to any claim at issue. *Id.* at 6 (quoting DOC 00314772 and DOC 02401762). Courts can and should

[2] Anthropic suggests regular expressions can be improved by "identifying" and removing "false positives." Opp. 4. But that is not what Dr. Bray did. He admits he retained all false positives, evidently to argue they offset the unknown number of false negatives. Dep. Tr. 49:8–13.

[3] Anthropic baselessly claims that "Publishers' logic" demands a "statistician-musicologist" to perform this analysis. Opp. 7. The parties have retained linguists like Dr. Leonard and musicologists like Dr. Lustig, and nothing prevented Dr. Bray from consulting other experts.

exclude opinions like this "that suffer[] from serious methodological flaws." *Teradata Corp. v. SAP SE*, 124 F.4th 555, 566 (9th Cir. 2024) (cleaned up and citations omitted). Anthropic cannot dismiss the dozens of classification errors Publishers identified as merely going to weight.

While Anthropic baldly asserts "courts regularly admit and credit expert testimony employing regex analysis," Opp. 4, it fails to identify a single such example. *Validity, Inc. v. Project Bordeaux, Inc.*, does not discuss expert testimony at all, only briefly mentioning regular expressions in describing the elements of a patent claim. 2023 WL 6200287, at *6 (D. Del. Sept. 22, 2023). Nor does the second case Anthropic relies on "approvingly cite[] an expert's use of regex analysis," as Anthropic claims. Opp. 4. In reality, that court chose not to rely on an expert's regex analysis and only found the plaintiffs met their burden on class certification through other means. *Norman v. Trans Union, LLC*, 669 F. Supp. 3d 351, 382 (E.D. Pa. 2023). In any event, the Court need not address whether regex analysis might ever be admissible; all that matters here is that Dr. Bray's error-ridden use of regex analysis to search for a general concept that admittedly can be expressed "[redacted]" is unreliable. Reply ¶ 34.

Anthropic's claim that Publishers' expert, Dr. Ben Zhao, agrees regular expressions "should be used" to "identify lyric-related ***prompts*** and ***outputs*** of an LLM," Opp. 4, mischaracterizes Dr. Zhao's opinion. Dr. Zhao addressed the use of regular expressions to "detect song ***lyrics***," not ***prompts*** requesting lyrics, in Anthropic's training corpus. ECF No. 750-4 ("Zhao Rpt."). This distinction is critical: while there are countless ways Anthropic's millions of users could request lyrics, there is only one way to sing "A Change is Gonna Come." Anthropic's mischaracterization of Dr. Zhao's opinion highlights the distinction it fails to grasp between searching for known, fixed texts like lyrics, where regular expressions are useful, and searching for examples of a general concept, where regular expressions are unreliable.

Finally, Anthropic argues Dr. Bray's regular expression is reliable because it identified 115 requests for lyrics submitted by Publishers' investigator. Opp. 8. First, that Dr. Bray's regular expression identified some true positives does not make his opinion reliable; even a stopped clock is right twice a day. Second, Anthropic ignores that Dr. Bray missed other requests for lyrics submitted by this same investigator. Add'l Decl. of Keith Howell in Supp. of Pls.' Reply ("Add'l

Howell Decl."), Ex. A (ANTH_SAMPLE000100979). Third, these investigator prompts were known to Dr. Bray when he designed his regular expressions, and capturing them says nothing about his ability to capture ***unknown*** prompts.

II. Anthropic admits Dr. Bray failed to calculate an error rate and designed his regular expressions for use in litigation, which both favor exclusion.

Aware of the "many" mistakes in his analysis, Dr. Bray chose not to calculate an error rate. Mot. 5–8. Anthropic argues this failure is not a *per se* bar, but it concedes the failure to calculate an error rate is one factor courts consider under *Daubert*. Opp. 8–9. That Dr. Bray could have calculated an error rate, but chose not to, weighs in favor of exclusion. *Allen v. Am. Cap. Ltd.*, 287 F. Supp. 3d 763, 787 (D. Ariz. 2017) (excluding expert's analysis because "there is no known rate of error despite the availability of statistical analysis"); *United States v. Saelee*, 162 F. Supp. 2d 1097, 1103 (D. Alaska 2001) (granting motion to exclude expert's analysis where "[t]here is little known about the error rates"). Anthropic also admits Dr. Bray's regex patterns were "of course" designed "for the purpose of litigation." Opp. at 7–8. While not dispositive, this too weighs in favor of exclusion and can be "very significant" where, as here, an expert's analysis lacks peer reviewed support. *Daubert v. Merrell Dow Pharm.*, 43 F.3d 1311, 1317 (9th Cir. 1995).

III. Anthropic essentially abandons Dr. Bray's unsupported "[redacted]" opinion.

Instead of actually calculating an error rate, Dr. Bray claims his regular expressions "[redacted]" false positives and false negatives. But he performed no analysis to support this opinion. Dr. Bray could have easily reviewed the [redacted] prompts he classified as requests for lyrics to determine how many contained actual requests for lyrics. Reply ¶ 63. As Anthropic concedes, he chose not to. Opp. 8. Anthropic also claims false negatives are "impossib[le]" to calculate because there is "no measure of what a 'ground truth would be.'" Opp. at 8–9 (quoting ECF No. 755-3 64:6-15). Yet, somehow, Dr. Bray opines that a number he did not calculate (*i.e.*, false positives) perfectly balances a number Anthropic claims is impossible to calculate (*i.e.*, false negatives). Unsurprisingly, this analysis is supported by nothing more than Dr. Bray's say-so and should be excluded. *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146 (1997) (courts should exclude "evidence that is connected to existing data only by the *ipse dixit* of the expert.").

Unable to coherently defend Dr. Bray's analysis, Anthropic claims that his regular expressions must balance false positives and false negatives because Publishers pointed out equal numbers of both. Opp. 8–9 (citing Mot. 6–8). But, as Publishers noted, their Motion includes only "a small sample of the errors in Dr. Bray's results," and nothing suggests false positives and false negatives are of an equal magnitude. This wholly unsupported opinion should be excluded.

IV. Dr. Bray's refusal to acknowledge copying after [redacted] collapses infringement and fair use inquiries and will mislead the jury.

Dr. Bray originally opined that "there are ***no outputs*** providing lyrics to the Works-in-Suit after [redacted]," simply because his word-search analysis for Question 4 produced no hits after that date. Rpt. ¶ 40 (emphasis added), ECF No. 656-2. Publishers' expert Mr. Dennis pointed out two examples that Dr. Bray missed. Add'l Howell Decl., Ex. B (Dennis Rebuttal) ¶ 59. Even though Dr. Bray acknowledged these outputs contain "the same sequence of words" as Publishers' "Genuine Lyrics," he stubbornly stuck with his original opinion by creating new distinctions not included in his original report and not based on any legal definition of copying. Dep. Tr. 180:13–181:11. He opined that these outputs "[redacted]" the parameters of records "[redacted]" because the "[redacted]" and Claude "[redacted]." Reply at ¶¶ 53–54. This departs from the only task for which Anthropic claims Dr. Bray is qualified: determining if a record "provid[es] lyrics to the Works-in-Suit." Anthropic's argument that Dr. Bray does not offer legal conclusions does not make it so: there is no other way to understand his refusal to acknowledge the existence of records he agrees contain "the same sequence of words" as Publishers' "Genuine Lyrics." Dep. Tr. 180:13–181:11. Nor is this opinion "squarely a factual statistical presentation," as Anthropic claims; there is nothing "statistical" about discounting copying of the Works in Suit because the copying is done to "[redacted]," or "[redacted]" Reply at ¶¶ 53–54.

CONCLUSION

The Court should exclude Dr. Bray's challenged opinions for the foregoing reasons.

Dated: August 18, 2026

Respectfully submitted,

/s/ Keith Howell

OPPENHEIM + ZEBRAK, LLP
Matthew J. Oppenheim
Nicholas C. Hailey
Jeffrey M. Gould
Corey Miller
Keith Howell
(admitted *pro hac vice*)
4530 Wisconsin Ave., NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
matt@oandzlaw.com
nick@oandzlaw.com
jeff@oandzlaw.com
corey@oandzlaw.com
khowell@oandzlaw.com

Alexander Kaplan
Jennifer Pariser
Andrew Guerra
Bret Matera
Timothy Chung
Michelle Gomez-Reichman
(admitted *pro hac vice*)
461 5th Avenue, 19th Floor
New York, NY 10017
Telephone: (212) 951-1156
alex@oandzlaw.com
jpariser@oandzlaw.com
andrew@oandzlaw.com
bmatera@oandzlaw.com
tchung@oandzlaw.com
mgomez-reichman@oandzlaw.com

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@cll.com
jzk@cll.com
rxd@cll.com

Attorneys for Plaintiffs

SIGNATURE ATTESTATION PURSUANT TO CIVIL L.R. 5-1(h)

Pursuant to Civil L.R. 5-1(h), I hereby attest that concurrence in the filing of this document was obtained from the signatory of this document. I declare under penalty of perjury that the foregoing is true and correct.

Dated: August 18, 2026

/s/ Timothy Chung
Timothy Chung