EXHIBIT G

HC-AEO

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

CONCORD MUSIC GROUP, INC., et al., Plaintiffs, v. ANTHROPIC PBC, Defendant.	Case No. 5:24-cv-03811-EKL-SVK

EXPERT REPLY REPORT OF ABHISHEK NAGARAJ, PH.D.

Signed in Tokyo, Japan on February 13, 2026



Abhishek Nagaraj, Ph.D.

Table of Contents

I. INTRODUCTION 4

A. QUALIFICATIONS 4

B. ASSIGNMENT 4

C. SUMMARY OF OPINIONS 4

II. IN DISCUSSING THE MARKETS FOR THE WORKS-IN-SUIT, THE SMITH REBUTTAL REPORT FAILS TO CORRECTLY ANALYZE THE RELEVANT BUYERS AND SELLERS 7

A. THE MARKET FOR THE DISPLAY OF LYRICS IS SEPARATE FROM ANY PURPORTED MARKET FOR LYRICS AS TRAINING DATA AND MUST BE ANALYZED SEPARATELY 9

B. DR. SMITH'S CRITIQUE THAT THE RELEVANT MARKET FOR AI TRAINING SHOULD CONSIDER LICENSES "FOR ANY PURPOSE" CONTRADICTS BASIC ECONOMIC PRINCIPLES 11

1. Dr. Smith Improperly Includes Irrelevant Buyers in His Analysis of a Market for Training Data 16

2. Dr. Smith Fails to Recognize That the Relevant Market for Training Data Contains a Broad Set of Potential Suppliers 21

C. DR. SMITH'S DISCUSSION OF TRANSACTION COSTS IS ERRONEOUS 24

1. Dr. Smith's Overall Opinion on Transaction Costs Is Flawed 24

2. Dr. Smith's Conclusion on Transaction Costs Regarding the Licensing of Lyrics is Flawed 27

D. THE SMITH REBUTTAL REPORT DOES NOT ACCURATELY ASSESS THE RELEVANCE OF EXISTING DEALS TO THE MARKETS IN QUESTION 35

E. DR. SMITH MISINTERPRETS MY PAPER "CLOZE ENCOUNTERS" 44

F. HARMS TO OTHER MARKETS THE SMITH REBUTTAL REPORT ADDRESS ARE IRRELEVANT 52

III. DR. SMITH'S CRITICISMS OF MY REGRESSION MODELS ARE FLAWED AND DO NOT CHANGE MY OPINIONS 54

A. THE ANALYSES I PRESENTED IN MY OPENING REPORT ADEQUATELY CONTROL FOR DEMAND SHOCKS 55

1. The Traffic Analysis I Presented in my Opening Report Adequately Controls for Demand Shocks .. 55

2. The Google Search Analysis Presented in my Opening Report Also Adequately Controls for Demand Shocks 61

B. MY GOOGLE SEARCH ANALYSIS IS ACCURATE AND INCLUDES AN APPROPRIATE CONTROL GROUP 65

C. DR. SMITH'S CRITIQUE REGARDING THE USE OF CLAUDE.AI TRAFFIC AS THE INDEPENDENT VARIABLE IN MY TRAFFIC ANALYSIS AND GOOGLE TRENDS ANALYSIS LACKS ANY MERIT 72

D. DR. SMITH'S CRITIQUE OF MY USE OF FIXED EFFECTS IS MISCONSTRUED AND, MORE IMPORTANTLY, IRRELEVANT TO THE RESULTS OF MY ANALYSIS 75

1. The Usage of Fixed Effects in Website Traffic Analysis Follows Standard Academic Practice 76

2. The Usage of Fixed Effects in the Google Search Analysis Is Appropriate, and Results Are Robust to Non-Inclusion 82

IV. SUPPLEMENTAL STATEMENT TO MY OPENING REPORT 84

V. CONCLUSION 84

VI. APPENDIX 86

A. DETAILS ON THE TRAFFIC ANALYSIS 86

1. Regression Results with Usage of Fixed Effects 86

2. Regression Results with Scaling the Claude.ai Traffic 89

3. Regression Results Accounting for Volatility of Websites 90

4. Regression Results with Aggregating Lyrics Traffic 91

B. DETAILS ON THE GOOGLE SEARCH ANALYSIS 93

1. Regression Results with Alternative Thresholds for Viral Songs 93

2. Methodology for Estimating Residuals of Song Index Used in Alternate Regression Estimates of Google Trends Analysis 97

3. Regression Results without Using Fixed Effects 99

C. LIST OF MATERIALS RELIED UPON 102

I. INTRODUCTION

A. Qualifications

1. I previously provided my qualifications in Section I.A of my Opening Report dated on December 19, 2025. My CV in Appendix A of my Rebuttal Report is unchanged.

B. Assignment

2. I have been asked by Counsel for Anthropic PBC ("Anthropic") to analyze economic issues related to fair use factor 4, which is "the effect of the use upon the potential market for or value of the copyrighted work."[1] As a part of this assignment, I submitted my Opening Report on December 19, 2025. I understand that Plaintiffs' experts also submitted their Opening Reports on December 19, 2025. I was then asked by Counsel to evaluate the opinions and conclusions made by Dr. Smith, as well as certain opinions made by Dr. Ben Zhao and Mr. Joshua Dennis. As a part of this assignment, I submitted my Rebuttal Report on January 23, 2026.

3. I understand that on January 23, 2026, Plaintiffs' expert Dr. Michael Smith also submitted his Rebuttal Report (the "Smith Rebuttal Report"), which reviewed and responded to sections of my Opening Report.[2] For this present report, I have been asked by Counsel to also evaluate the opinions and conclusions made by Dr. Smith in his Rebuttal Report.

C. Summary of Opinions

4. Dr. Smith contends that my Opening Report erred[3] in assessing three actual or potential markets to determine "the effect of the use upon the potential market for or value of the copyrighted work"[4] of the Plaintiffs' Works-in-Suit. In fact, Dr. Smith's own market definition is flawed. Dr. Smith incorrectly describes the good at issue in this case as the right to "use the copyrighted works for ***any purpose***;"[5] he also includes in his market

[1] Opening Report of Abhishek Nagaraj, 19 Dec 2025 ("Nagaraj Opening Report"), at ¶6.

[2] *See* Rebuttal Report of Michael D. Smith, 23 Jan 2026 ("Smith Rebuttal Report"), at ¶2.

[3] *See* Smith Rebuttal Report, at ¶¶9-14.

[4] *See* 17 U.S. Code § 107, "Limitations on exclusive rights: Fair use" ("In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include…the effect of the use upon the potential market for or value of the copyrighted work.").

[5] Smith Rebuttal Report, at ¶15 (emphasis added).

lyrics. Those hypothetical substitutes plausibly consist of both lyrics[94] and non-lyrics text data, corresponding to the information I discuss in my first purported market.[95]

46. In fact, in my Opening Report I explained that "song lyrics have low value for training general-purpose LLMs because there are other ways to help train LLMs to answer user prompts regarding songs and songwriting."[96] Dr. Bilmes also points out that, to the extent it is useful for a general-purpose LLM to understand music, lyrics, and song structure, this information is also available from other sources, including books and internet text.[97] Thus, there are many other substitutes for lyrics data from Plaintiffs' copyrighted works, which Dr. Smith fails to consider.

47. In sum, Dr. Smith ignores the fact that the relevant sellers are those with viable substitutes for Plaintiffs' Works-in-Suit as training data. In fact, even if the relevant market were limited to lyrics text data, it would necessarily include all sources of lyrics text data (such as lyrics to songs owned by other music labels), not solely Plaintiffs' works. As discussed throughout my Opening Report and Rebuttal Report,[98] developers of general-purpose LLMs need to build a vast and diverse training corpus through a mix of textual data originating from different sources or providers, not only Plaintiffs. At the same time, including data from two content suppliers with significantly overlapping information provides little additional value. Lyrics in particular are not a unique or indispensable input, as their informational content can be replaced by other textual sources. Accordingly, Dr. Smith errs in overlooking the role of substitutes to Plaintiffs' Works-in-Suit and fails to recognize that Anthropic could substitute Plaintiffs' Works-in-Suit

---

[94] For example, lyrics from other publishers and songwriters could substitute for the Works-in-Suit, including lyrics in the public domain.

[95] *See* Nagaraj Opening Report, at Section III.A.2.

[96] *See* Nagaraj Opening Report, at ¶60.

[97] *See* Bilmes Opening Report, at ¶330 ("To the extent that it is helpful for Claude to understand music, lyrics, and the structure of songs as part of Anthropic's goal of creating a versatile, general-purpose model, there are ample data sources in Claude's training corpus—other than the song lyrics themselves—where this information is available. These alternative sources, such as books, the internet, and open-domain lyrics, would suffice to enable Claude to understand those concepts.").

[98] *See* Nagaraj Opening Report, at Section II.B; Nagaraj Rebuttal Report, at ¶102 ("As I stated in my Opening Report, in a purported market for all of the data needed to train a general-purpose LLM, the transaction costs would dwarf the value of the data, because the 'data needed to train a general-purpose LLM is immense and incredibly diverse, and the potential holders of rights in that data are enormous and highly distributed.'"); Nagaraj Rebuttal Report, at ¶121 ("LLM developers like Anthropic are likely to place a low value on lyrics given how small a lyrics dataset would be in comparison to the size needed to train a general- purpose LLM. . . . [Song lyrics] tend to include a far narrower range of topics, which minimizes the diversity that they contribute. For this reason, they add little to the range and diversity of knowledge that a general-purpose LLM like Claude needs."); Nagaraj Rebuttal Report, at footnote 188 ("Ironically, the larger and more diverse that the dataset on which a generative model was trained is, the less likely are infringing outputs."). Citations are omitted.

with other text sources—including text that does not consist of lyrics—to train its general-purpose LLM.

C. Dr. Smith's Discussion of Transaction Costs Is Erroneous

1. Dr. Smith's Overall Opinion on Transaction Costs Is Flawed

48. In his Rebuttal Report, Dr. Smith critiques the market I consider to be most relevant to training data, which is the purported market for all the data needed to train a general-purpose LLM,[99] and claims that my opinion about the high transaction costs in such a market "collapses into circularity."[100] He further notes that a "general-purpose LLM developer, like Anthropic, could have decided to license copyrighted content first instead of obtaining it without licensing it. Had any of the buyers in Dr. Nagaraj's market chosen to do so, then, by definition, a market for licensing that content would exist."[101] But Dr. Smith fails to grapple with my conclusion that, assuming the relevant market is for all data needed to train a general-purpose LLM, the transaction costs of acquiring the trillions of tokens needed would be insurmountable and likely prevent a market from forming.[102]

49. His contention ignores the many factors I point to in my Opening Report that contribute to these high transaction costs: the vast scale of the data needed,[103] the diversity of the data required,[104] the large number of rightsholders involved,[105] the difficulty of locating

---

[99] *See* Smith Rebuttal Report, at ¶8 ("Dr. Nagaraj errs by focusing the bulk of his 'factor four' analysis on two markets-neither of which bear on the 'potential market for and value of' ***Publishers' lyrics***, the copyrighted works at issue-and both of which lead him to misleadingly conclude that 'there can have been no harm to Plaintiff s in ***these*** (nonexistent) markets.'") (emphasis in original); *see also* Smith Rebuttal Report, at Section II.A.

[100] Smith Rebuttal Report, at ¶39.

[101] Smith Rebuttal Report, at ¶41.

[102] *See* Smith Rebuttal Report, at ¶42 ("Dr. Nagaraj's opinions regarding 'prohibitively high transaction costs' are irrelevant to his stated assignment. Whether transaction costs would prohibit a specific class of developer from assembling 'All of the Data Needed' to train a specific type of AI model does not mean that a 'potential market" for Publishers' lyrics does not or cannot exist . . . .").

[103] *See, e.g.,* Nagaraj Opening Report, at Section II.B ("Training an LLM requires trillions of tokens to allow the model to recognize the patterns and relationships between words and phrases, and to eventually develop a comprehensive map of how language works."); Nagaraj Opening Report, at ¶¶36-37 ("According to Anthropic's Chief Science Officer, Dr. Kaplan, LLMs like Claude must be 'shown a vast dataset consisting of hundreds of millions, if not billions, of pieces of content.' . . . [redacted] and Nagaraj Opening Report, at ¶47 ("[E]even the largest rightsholders represent only a fraction of the data needed to train a general-purpose LLM such as Claude. . . . One of the reasons the [OpenAI and *The New York Times*] deal fell through, according to OpenAI, was that even a highly regarded corpus like the one belonging to *The New York Times* did not represent a large enough dataset relative to OpenAI's overall data needs.").

[104] *See* Nagaraj Opening Report, at ¶38 ("As Dr. Kaplan stated, 'any functional general-purpose LLM requires a very diverse set of training data to gain a sufficiently precise understanding of the full range of human language and human experience.' Without using a variety of sources for training datasets, developers find that their models degrade.").

[105] *See* Nagaraj Opening Report, at ¶42 ("[B]ecause of the scale of training data required for general-purpose LLMs, developers would need to make deals with hundreds of millions of different rightsholders.").