EXHIBIT 2

(FILED UNDER SEAL)

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL.,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

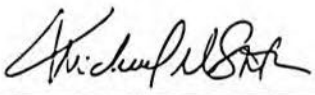
Case No. 5:24-cv-03811-EKL-SVK

DESIGNATED HIGHLY CONFIDENTIAL ATTORNEYS' EYES ONLY

Judge Eumi K. Lee
Magistrate Judge Susan van Keulen

REBUTTAL EXPERT OPINION OF PROFESSOR MICHAEL D. SMITH

Dated: January 23, 2026



Michael D. Smith

Table of Contents

I. OVERVIEW AND ASSIGNMENT 3

II. REBUTTAL TO DR. NAGARAJ'S OPINIONS 3

A. DR. NAGARAJ'S "FACTOR FOUR" ANALYSIS OF A MARKET TO "ASSEMBLE ALL OF THE DATA NEEDED TO TRAIN A GENERAL-PURPOSE LLM" IS FLAWED. 5

1. Summary of opinions 5

2. Dr. Nagaraj excludes all but a narrow class of buy-side market participants from his "four factor" analysis. 7

3. Dr. Nagaraj skews his factor four analysis by framing the market he analyzes in terms of a buyer's "needs" to train a general-purpose LLM. 10

4. Dr. Nagaraj's opinion regarding prohibitive transaction costs is irrelevant and circular 19

5. Dr. Nagaraj appears to further limit his analysis to the year 2021. 30

B. DR. NAGARAJ'S "FACTOR FOUR" ANALYSIS OF A MARKET FOR LYRICS USED TO TRAIN A GENERAL-PURPOSE LLM IS FLAWED. 31

1. Dr. Nagaraj uses the same constrained and unreliable market parameters regarding General-Purpose LLM and buyer "needs." 32

2. Dr. Nagaraj unnecessarily frames his alternative market in terms of General-Purpose LLM developers attempting to train their models with only song lyrics. 33

3. Dr. Nagaraj's opinions regarding the valuation of music lyrics rest on erroneous and conflicting assumptions. 33

4. Dr. Nagaraj's opinions about transaction costs for licensing music lyrics are speculative and irrelevant. 39

C. DR. NAGARAJ'S OPINIONS REGARDING THE RELEVANCE OF EXISTING DEALS TO LICENSE AI TRAINING DATA CONTRADICT HIS MARKET OPINIONS 41

D. DR. NAGARAJ IGNORES THE EFFECT OF ANTHROPIC'S USE ON OTHER RELEVANT MARKETS 44

E. DR. NAGARAJ'S REGRESSION ANALYSES CONTAIN NUMEROUS FLAWS AND OBSCURE THE RELATIONSHIP BETWEEN CLAUDE AND LYRIC INTEREST. 47

1. As designed, Dr. Nagaraj's regression models are incapable of evaluating the causality that he purports to analyze. 48

2. The variables of interest in Dr. Nagaraj's models are manipulated and do not reflect the conduct at issue. 55

3. Dr. Nagaraj's use of fixed effects obscures the low explanatory power of his models and the relationship between Claude and lyric interest. 60

III. REBUTTAL TO MS. DAVIS'S OPINIONS 65

REBUTTAL EXHIBIT A I

APPENDIX 1: DOCUMENTS CONSIDERED V

I. OVERVIEW AND ASSIGNMENT

1. On December 19, 2025, I issued an initial expert report in this case. *See* Expert Report of Michael D. Smith (Dec. 19, 2025) (hereinafter the "Smith Report"). On the same day that I served my initial report, Defendant Anthropic PBC ("Anthropic") served a series of expert reports from Dr. Abhishek Nagaraj, Julie L. Davis, Dr. Andrew Bray, Collin Stokes, Dr. Jeff Bilmes, and Russ B. Altman.

2. As part of my retention by Publishers, I have been asked to review and comment on sections of the expert reports of Dr. Nagaraj (*see* Expert Report of Dr. Abhishek Nagaraj (Dec. 19, 2025), hereinafter the "Nagaraj Report") and Ms. Davis (*see* Expert Report of Julie L. Davis (Dec. 19, 2025), hereinafter the "Davis Report"), that are relevant to my expert work in this matter. In this rebuttal expert report, I provide my assessment of and opinions about certain aspects of the expert work performed by Anthropic's experts. The materials that I considered in preparing my expert reports, including this rebuttal, are cited herein and/or included in Appendix 1.

3. My credentials and compensation are set forth in the Smith Report, which I incorporate herein by reference. My compensation is not contingent or dependent on the assessment and opinions in this rebuttal report nor the outcome of this or any other matter

4. To the extent that Anthropic serves additional expert reports or supplements the reports it has already served, I reserve the right to supplement my analyses and opinions.[1]

II. REBUTTAL TO DR. NAGARAJ'S OPINIONS

5. Dr. Nagaraj states that he was engaged "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

[1] I continue to reserve the right to supplement or modify my analyses based on recent or forthcoming discovery provided by Anthropic, including newly discovered information as it becomes available.

work'" (referred to herein as Dr. Nagaraj's "stated assignment"). Nagaraj Report ¶ 6. Specifically, he states that he was "asked to assess this factor in relation to Plaintiffs' claims regarding Anthropic's alleged use of lyrics from songs to which Plaintiffs hold a copyright." *Id.* He also states that he was "asked to assess fair use factor 4 in relation to Claude's outputs and the purported recitation of partial or complete lyrics in songs to which Plaintiffs hold a copyright." *Id.*

6. In furtherance of his stated assignment, Dr. Nagaraj purports to analyze "whether Anthropic's alleged use of lyrics from the Works-In-Suit to train commercial models of Claude and Claude's alleged display of those lyrics as outputs had an effect on three purported markets": (1) "a purported market for all of the data needed to train a general-purpose LLM"; (2) "a purported market for song lyrics to train a general-purpose LLM"; and (3) "a market for the online display of lyrics." *Id.* ¶¶ 25–26.

7. For the purpose of this rebuttal report, I focus on Dr. Nagaraj's "factor four analysis" opinions related to the "effect of [Anthropic's] use on the potential market for and value of" Publishers' copyrighted works. *Id.* ¶ 6.

8. Based on my review of the content of his report, the evidence he cites, his own prior publications, and empirical standards in economic literature, I find that Dr. Nagaraj's opinions and analysis rest on fundamentally flawed foundations that cause him to make conclusions that are misleading, unreliable, and irrelevant to his own stated assignment. 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 Plaintiffs in ***these*** (nonexistent) markets." *Id.* ¶ 27 (emphasis added). Similarly, his analysis related to "a market for

the online display of lyrics" also contains numerous methodological flaws, which I describe in detail below. *See* below § II.E.

A. Dr. Nagaraj's "Factor four" Analysis of a Market to "Assemble All of the Data Needed to Train a General-Purpose LLM" is Flawed.

1. Summary of opinions.

9. As noted above, Dr. Nagaraj's stated assignment was "to analyze economic issues related to fair use factor 4, which is 'the effect of [Anthropic's] use upon the potential market for or value of the copyrighted work'" at issue. Nagaraj Report ¶ 6. But the principal "market" that Dr. Nagaraj claims is "the correct market to analyze" for this assignment inherently skews his subsequent analysis for at least four reasons. *Id.* ¶ 8. It should also be noted at the outset that another Anthropic expert, Julie Davis, whose opinions I discuss below in § III, disagrees with Dr. Nagaraj about what constitutes the "correct" market for their respective "factor four" analyses.[2]

10. First, Dr. Nagaraj narrowly defines the relevant market to include only a small "buyer" class of "companies like Anthropic that are building general-purpose LLMs," while excluding all other potential "buyers" of the copyrighted works at issue. Nagaraj Report ¶ 31. Such a narrow definition of "buyer" inherently limits and skews Dr. Nagaraj's analysis because he does not even consider what effect Anthropic's use has had on other potential "buyers" of the works at issue, including other companies that are building "single-purpose" AI models (as opposed to what Dr. Nagaraj calls "General-Purpose LLMs").

11. Second, Dr. Nagaraj frames this market in terms of what a specific class of buyer "needs" to achieve a specific (and subjective) business objective. Specifically, he frames the market in terms of "All of the Data ***Needed to Train*** a General-Purpose LLM." Nagaraj Report

---

[2] Specifically, Ms. Davis defines the relevant for her "factor four" analysis as simply the "market for licensing the lyrics of the Asserted Works." *See* Davis Report at p. 3.

¶ 33 (emphasis added). This framing conflicts with Dr. Nagaraj's own stated assignment of assessing "the effect of [Anthropic's] use upon the ***potential market for or value of the copyrighted work***," *id.* ¶ 6 (emphasis added), *i.e.*, the potential market for "***selling***" the copyrighted work. By framing his market model this way, Dr. Nagaraj inherently divorces his subsequent analysis from his stated assignment.

12. Third, Dr. Nagaraj's conclusion regarding "prohibitively large transaction costs" rests on a flawed assumption about the valuation of copyrighted works that is contradicted by the facts in the case, Anthropic's party admissions, Dr. Nagaraj's prior publications, and statements contained in Dr. Nagaraj's own report. Additionally, despite his stated assignment to assess the market for the copyrighted works at issue, Dr. Nagaraj devotes much of his transaction-cost analysis to assessing the feasibility of licensing vast numbers of irrelevant "orphan" works that are unrelated to this case.

13. Last, Dr. Nagaraj appears to conclude that a licensing market for Publishers' copyrighted works could not have existed in 2021 and is not likely to in the future, *see* Nagaraj Report ¶ 53, even though Anthropic was not founded until 2021 and the first Generative AI model was not publicly released until 2022. Historical evidence shows, and Dr. Nagaraj even admits in his report, Nagaraj Report ¶ 67, that licensing frameworks for new technologies take time to develop and establish. The fact that a fully-formed licensing market does not immediately emerge within the first few years of a technology's creation does not mean that such market did not exist at all—and certainly does not support a conclusion that one is not likely to *ever* develop in the future.

14. As I discuss below, for these reasons, and more, Dr. Nagaraj's analysis of a "Market to Assemble All of the Data Needed to Train a General-Purpose LLM" is inherently skewed and

unreliable in addressing his own stated assignment of assessing the "effect of [Anthropic's] use upon the potential market for or value of the copyrighted work[s]" at issue in this case.

2. Dr. Nagaraj excludes all but a narrow class of buy-side market participants from his "four factor" analysis.

15. Dr. Nagaraj limits his definition of relevant buyers in his "four factor" analysis to only those "companies like Anthropic that are building general-purpose LLMs." Nagaraj Report ¶ 31. This definition excludes all classes of potential buyers of the at-issue works except for those very few who will use the copyrighted works specifically for training "general-purpose LLMs." By defining the market this way, Dr. Nagaraj excludes all other buyers who would use the copyrighted works for any purpose, including to train other types of AI models (such as "single-purpose" AI models, as opposed to "general-purpose" models). In the context of Dr. Nagaraj's stated assignment to analyze "the effect of [Anthropic's] use upon the potential market for or value of the copyrighted work," this definition of buyers is unnecessarily restrictive and limits the relevance and reliability of Dr. Nagaraj's subsequent analysis.

16. As I discuss in my initial report, there are many prospective buyers of Publishers' works for a variety of uses, including for use as AI training data. For example, two prominent examples that I highlighted in my initial report were the deals that UMPG struck with Uncharted Labs, Inc. (d/b/a Udio.com, "Udio") and Klay Vision, Inc. ("Klay"). *See* Smith Report ¶ 120. In both deals, UMPG licensed the rights to its catalog of music compositions, including lyrics, for use in training Udio's and Klay's respective AI models. UMPG_00000024457, at -60 [redacted] UMPG_0000025119, at -24 [redacted]). Similarly, in my initial report, I also highlighted Publishers' various (and lucrative) deals with digital distribution platforms, including deals with Apple, Google, Spotify, and Meta,

see Smith Report ¶ 163, and with online lyric-display websites like LyricFind, MusixMatch, and Genius, Smith Report ¶ 164. Each of these entities (and their respective licensing deals with Publishers) help make up the "potential market for and value of the copyrighted work[s]" at issue here, and yet Dr. Nagaraj excludes them from his definition of buyer (and, by consequence, from his "factor four" analysis).

17. Instead, Dr. Nagaraj narrowly defines the relevant buyers to include only "companies like Anthropic that are building general-purpose LLMs." Nagaraj Report ¶ 31. He defines "general-purpose LLM" as a text-based generative AI model "built to perform a wide variety of tasks such as understanding, summarizing, predicting, and generating new content based on vast and diverse data." *Id.* ¶ 20. While Dr. Nagaraj never identifies all of the buyers in this market, as far as I am aware, there are only four companies in the United States building AI models that fit Dr. Nagaraj's definition of "general-purpose LLM," and therefore the market he chooses to analyze includes only four potential buyers: Anthropic (Claude), Meta (Llama),[3] Google (Gemini),[4] and OpenAI (ChatGPT).[5] Notably, all four of these companies are currently being sued or have faced lawsuits for copyright infringement in connection with their unlicensed use of copyrighted content to train their respective AI models.[6]

---

[3] *See* META, *Llama 4: Leading Intelligence*, https://www.llama.com/ (last visited Jan. 22, 2026).

[4] *See* GOOGLE, *Gemini*, https://gemini.google.com (last visited Jan. 22, 2026).

[5] *See* OPENAI, *ChatGPT*, https://chatgpt.com/overview/ (last visited Jan. 22, 2026).

[6] *See, e.g.*, *Kadrey et al. v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC (N.D. Cal.); *In re Google Generative AI Copyright Litigation*, No. 5:23-cv-03440-EKL (N.D. Cal); *Huckabee v. Meta Platforms, Inc.*, No. 24-cv-773-GBW (D. Del.); *DENIAL v. OpenAI, Inc.*, 3:25-cv-05495 (N.D. Cal.); *see also In re: OpenAI, Inc. Copyright Infringement Litigation*, No. 1:25-md-3143 (S.D. Cal.) (consisting of: *Tremblay v. OpenAI, Inc.*, No. 3:23-cv-03223-AMO (N.D. Cal.); *Silverman v. OpenAI, Inc.*, No. 3:23-cv-03416 (N.D. Cal.); *Chabon v. OpenAI, Inc.*, No. 3:23-cv-04625 (N.D. Cal.); *Millette v. OpenAI, Inc.*, No. 5:24-cv-04710 (N.D. Cal.); *Authors Guild v. OpenAI, Inc.*, No. 1:23-cv-08292 (N.D. Cal.); *Alter v. OpenAI, Inc.*, No. 1:23-cv-10211-SHS (N.D. Cal.); *The New*

18. Even if Dr. Nagaraj wanted to focus his analysis only on the market for Publishers' works as AI training data, specifically, he should have included other types of AI companies in his analysis, including what he refers to in his report as "single-purpose generative AI companies like Klay, Wondera, and ElevenLabs." Nagaraj Report ¶ 78. The standard economic definition of a market includes ***all*** actual and potential buyers of a product. For example, Pindyck and Rubinfeld define a "market" as a "[c]ollection of buyers and sellers that, through their actual or potential interactions, determine the price of a product or set of products."[7] As such, following standard economic definitions, to address his stated assignment, Dr. Nagaraj should have included single-purpose AI companies in his market analysis. Such single-purpose AI companies are potential buyers of Publishers lyrics for use as AI training data whose interactions determine the price for that use and who contribute to the "potential market for and value of" Publishers' copyrighted works. *See, e.g.*, Smith Report ¶ 120. Further, many use copyrighted text-based inputs for training, just like general-purpose LLMs.

19. Moreover, by Dr. Nagaraj's own definition, general-purpose LLMs necessarily encompass many (if not all) of the use cases performed by single-purpose AI models because general-purpose LLMs are "built to perform a wide variety of tasks such as understanding, summarizing, predicting, and generating new content based on vast and diverse data" and "need to be familiar with *everything*." Nagaraj Report ¶ 20 (emphasis added). For example, as I explained in my initial report, songwriting is one use case that Anthropic has enabled Claude to perform. *See* Smith Report ¶¶ 85–92, 201–02. With respect to this specific use case, Anthropic competes with

---

York Times v. OpenAI, Inc., et al., No. 1:23-cv-11195 (S.D.N.Y); *Raw Story Media, Inc. v. OpenAI, Inc.*, No. 1:24-cv-01514 (N.D. Cal.); *Intercept Media, Inc. v. OpenAI, Inc.*, No. 1:24-cv-01515-JSR (S.D.N.Y.); *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 1:24-cv-04872 (S.D.N.Y.)).

[7] Robert S. Pindyck & Daniel L. Rubinfeld, MICROECONOMICS, 8th ed. (2013).

single-purpose AI models designed for that specific task. Anthropic therefore necessarily competes with those single-purpose models to acquire whatever data is necessary to train an AI model to perform that task, which reasonably would include song lyrics. Anthropic and other AI models that are specifically designed for songwriting (like Udio and Klay) are therefore potential "buyers" in the same market for Publishers' lyrics and will use those lyrics for the same or substantially similar purpose in AI training. Thus, if Anthropic is able to train on song lyrics for free (by scraping the data or otherwise) while single-purpose models pay to train on those same lyrics through licensing, Anthropic obtains a competitive advantage for that use case vis-à-vis the single-purpose.

20. Thus, to properly assess the effect of Anthropic's alleged infringement "upon the potential market for and value of the copyrighted work[s]" at issue, as Dr. Nagaraj claims to be doing with his analysis, Dr. Nagaraj should have defined the buyers to at least encompass other companies seeking to use the works at issue for training single- and specific-use AI models, not just the select few companies that would use Publishers' works to train a general-purpose LLM. Had he done so, Dr. Nagaraj would have been unable to conclude that actual and potential markets for Publishers works do not exist given the growing number of licensing agreements between AI companies and rightsholders that are cited and discussed in my initial report. Smith Report ¶¶ 51–52 & Exhibit A.

3. Dr. Nagaraj skews his factor four analysis by framing the market he analyzes in terms of a buyer's "needs" to train a general-purpose LLM.

21. In addition to narrowing his definition of buyer to include only a select few general-purpose LLM developers, Dr. Nagaraj further distorts his factor four analysis by framing the market he analyzes in terms of "All of the Data ***Needed*** to Train a General-Purpose LLM." Nagaraj Report ¶¶ 8–9, 27, 33 (emphasis added). By framing his market this way, Dr. Nagaraj inherently skews his subsequent factor-four analysis.

22. First, by framing his market in terms of a specific buyer's "needs," Dr. Nagaraj puts his subsequent analysis in clear tension with his own stated assignment "to analyze economic issues related to fair use factor 4, which is 'the effect of [Anthropic's] use upon the potential market for or value ***of the copyrighted work***.'" Nagaraj Report ¶ 6 (quoting 17 U.S.C. § 107). By the terms of his assignment (and the statute that he quotes), the relevant market he should be analyzing is the market for ***Publishers' lyrics***. Indeed, this is exactly how Anthropic's expert Julie Davis defined the relevant market for her own "factor four" analysis. *See* Davis Report at p. 49 ("I consider whether the Plaintiffs' financial data support their claim that there has been harm to the market for licensing the lyrics ***of the Asserted Works***.") (emphasis added). Dr. Nagaraj's failure to properly frame his market in terms of the copyrighted works at issue undermines all of his subsequent factor-four analysis and opinions and leads him to make multiple misleading and irrelevant conclusions.

23. For example, Dr. Nagaraj misleadingly concludes that "no actual or potential market is likely to exist in which a developer can assemble all of the data it needs to train a general-purpose LLM." Nagaraj Report ¶ 35. But whether this purported market exists or not is immaterial to Dr. Nagaraj's stated assignment, which asks whether a market for ***Publishers' lyrics*** exists. Dr. Nagaraj thus ignores the possibility that a market for Publishers' lyrics as training data could exist even if a market for all of the data needed to train a general-purpose LLM did not exist. Nevertheless, Dr. Nagaraj asserts that his market framing is "correct. . . because LLM developers need to be able to amass a large enough set of training data in order to train a capable general-purpose LLM[.]" *Id.* ¶ 8. There are multiple problems with this justification.

24. For starters, Dr. Nagaraj does not explain what it means to train a "capable" general-purpose LLM, *id.* ¶ 8, nor does he explain the significance of this "capable" qualifier,

making this qualification an inherently subjective determination. Moreover, Dr. Nagaraj does not explain why a market cannot exist where LLM developers are only able to acquire *some* of the data needed to train a "capable" AI model, or even a less-than-capable AI model. This omission from Dr. Nagaraj's analysis is particularly glaring in light of his express acknowledgment that "occasional deals for training data might materialize," *id.* ¶ 33, and that "companies that develop LLMs could enter into deals to acquire data in specific circumstances," including deals for "gap-filling" and "last-mile data." *Id.* ¶¶ 72–73. Indeed, Dr. Nagaraj admits that [redacted] and I discussed many similar deals in my initial report. *See* Smith Report ¶ 148 (noting that Anthropic has purchased $66 million worth of print books through its data-acquisition department).

25. Dr. Nagaraj therefore admits that some markets for AI training data already exist, but he makes no attempt to analyze these markets or the effects of Anthropic's use on them, nor does he attempt to analyze how these markets relate to Publishers' lyrics (as required by his stated assignment). He even admits that he is "aware of the existence of certain agreements to license music for training data in certain LLMs," Nagaraj Report ¶ 74, which confirms that a market "to license music for training data" must already exist (although he does not cite any specific agreements). Dr. Nagaraj nevertheless maintains that this market cannot exist because such deals for music training data are not "commonplace and routine." *Id.* ¶ 74. But Dr. Nagaraj does not explain what he means by "commonplace and routine," nor does he explain why the frequency of such deals would be dispositive to determining whether such a market exists. He ignores that AI technology itself is only a few years old (indeed, as noted above, Anthropic itself was only founded

in 2021), and that, even in this young market, hundreds of AI licensing deals have already been publicly announced.

26. To illustrate just how fast the AI licensing market is growing, Figure 1 below reflects the cumulative total of all new, publicly announced AI licenses since October 2022, as reported by the Copyright Alliance, a trade organization representing the creative industry.[8] My methodology for creating this chart is explained in more detail in Rebuttal Exhibit A.

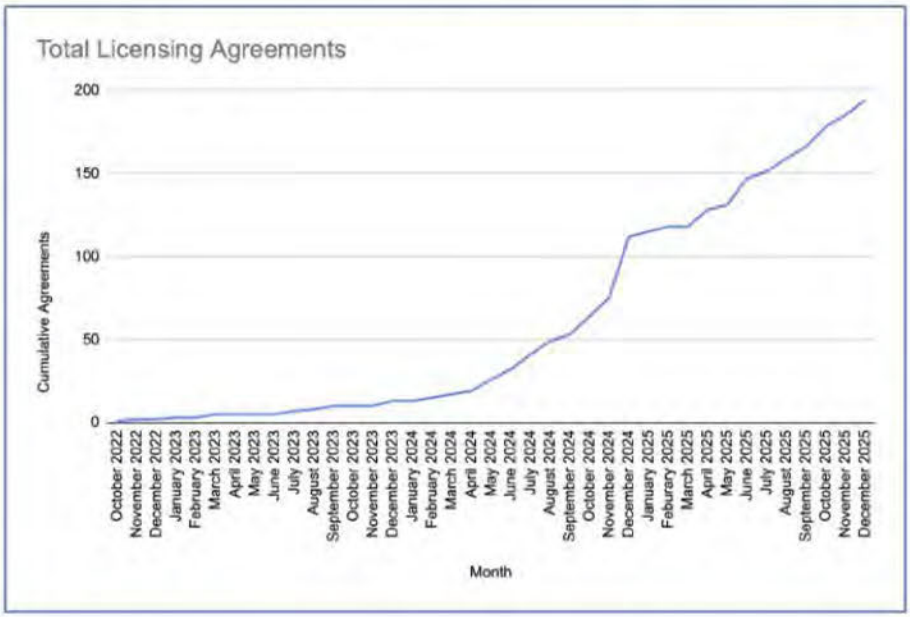


Figure 1

27. Figure 2 Below limits the data in Figure 1 to just licenses reported to be signed by one of the four general-purpose LLMs: Anthropic, Google, Meta, and OpenAI.

---

[8] *See* COPYRIGHT ALLIANCE, *AI Licensing by AI Companies*, https://copyrightalliance.org/artificial-intelligence-copyright/licensing/ai-companies/ (last visited Jan. 22, 2026).

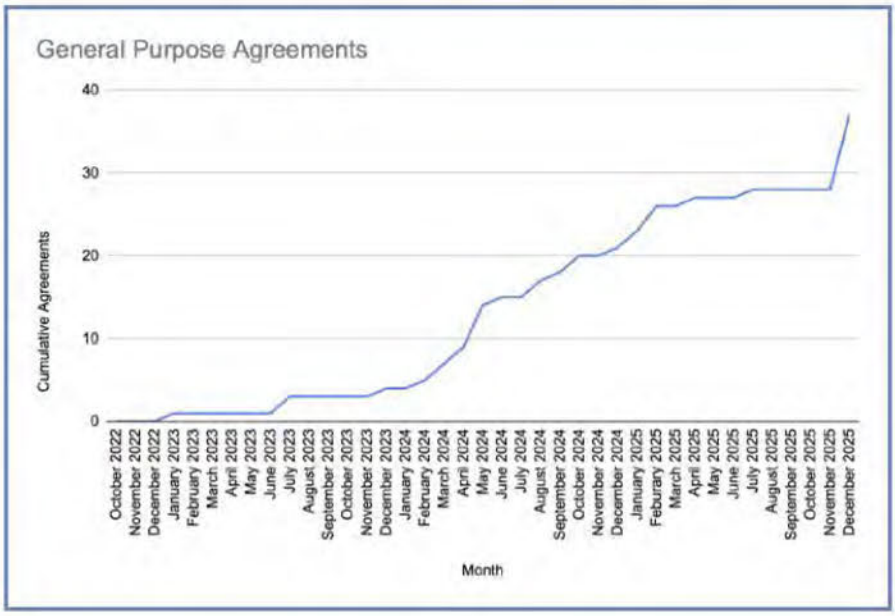


Figure 2

28. These figures demonstrate the existence of a robust AI licensing market that includes general-purpose LLMs and which it is growing rapidly, both in terms of all AI developers (growing from 13 deals in December 2023, to 112 deals in December 2024, to 194 in December 2025) and in terms of licenses with general-purpose LLM developers (growing from just four in December 2023, to 21 in December 2024, and 37 by December 2025).[9]

29. Second, by framing his market in terms of "***All of the Data*** Needed to Train a General-Purpose Model," Dr. Nagaraj introduces several additional non-economic considerations that further skew his subsequent analysis and lead him to additional misleading conclusions.

30. "All of the Data Needed" is an inherently flawed starting point for a market analysis because what a model "needs" is a subjective determination made by the model's developer. As Claude illustrates, not even Anthropic knows precisely how much data is "needed" to train its

[9] I also note, that of these 37 licenses with general-purpose LLMs, 22 are with OpenAI, 11 are with Meta, three are with Google, and only one is with Anthropic.

model. *See, e.g.*, P. Sangani 30(b)(1) Tr. 55:12–20. Moreover, the data that Anthropic uses to train its Claude models changes all the time, *id.*, making this metric—by definition—a moving target not suitable for empirical analysis. Other companies also likely disagree with Anthropic (and Dr. Nagaraj) about what is "needed" to train an AI model, and some companies have even publicly declared that it may be possible to fully train a "capable" LLM solely on open-sourced material.[10]

31. Dr. Nagaraj's "All of the Data Needed" framing also misleadingly assumes that a market for training data is an all-or-nothing proposition: either the model developer can license *all* of the data they need, or they will not license *any* data and choose not to develop any model at all. This assumption is unsupported and incorrect, including because it ignores the possibility that a developer will attempt to develop a model with a smaller dataset if they are unable to license or otherwise legally obtain *all* of the data that they subjectively believe would be ideal for their intended purposes (as would be the case in the [redacted] N[redacted] While training the model on a smaller dataset may affect performance, bigger is not always better; adding more data to the training set does not always lead to optimal model performance. *See* Expert Rebuttal Report of Benjamin Zhao (Jan. 23, 2026) (the "Zhao Rebuttal Report"), at ¶ 23; Expert Report of Dr. Jeff Bilmes (Dec. 19, 2025), at ¶ 165 & n.311. Dr. Nagaraj offers no reason to believe that an LLM developer would *not even attempt* to develop a model with less data, even if that means building one that is possibly less powerful than it would be with more data. As I noted above, other AI labs appear to be doing precisely that by using only licensed or public-domain content.[11]

---

[10] *See* Kate Knibbs, *Here's Proof You Can Train an AI Model Without Slurping Copyrighted Content*, WIRED (Mar. 20, 2024) (available at https://www.wired.com/story/proof-you-can-train-ai-without-slurping-copyrighted-content/) (last visited Jan. 23, 2026).

[11] *Id.*

32. Third, by framing his market in terms of "All Data Needed," Dr. Nagaraj bakes into his analysis an implicit and erroneous assumption that AI training data is fungible in its incremental value to improving model performance. Specifically, throughout his report and analysis, Dr. Nagaraj describes the "sheer scale and variety of the data needed to train LLMs." *See* Nagaraj Report ¶ 40. And he devotes many paragraphs of his report describing the vast data needs of general-purpose LLMs, *see, e.g.*, *id.* ¶¶ 34–37, which he then uses as justification to misleadingly conclude that "each piece of data on its own has very little incremental value to the training corpus." *Id.* ¶ 50. ("In other words, the additional value that a new piece of data contributes to an extremely large training corpus is minimal.").

33. Dr. Nagaraj's own prior statements and publications contradict this misleading conclusion. In their March 2025 article, *Cloze Encounters: The Impact of Pirated Data Access on LLM Performance*, Dr. Nagaraj and Stella Jia expressly "refute the claim that any single source has a negligible effect on model performance, *even in large production models*."[12] Specifically, Jia and Nagaraj find that leading LLM foundation models, including Claude, that were trained on a specific set of pirated works exhibited improved text-prediction performance as compared to foundation models that were not trained on those works. What's more, they find that this improved performance gets better when the models were trained on increasingly unique works (*i.e.*, less popular works that do not appear as frequently in online datasets), which, according to Jia and Nagaraj, "support[s] more bargaining power for content providers who provide more unique

[12] Stella Jia & Abhishek Nagaraj, *Cloze Encounters: The Impact of Pirated Data Access on LLM Performance*, NAT'L BUREAU OF ECON. R., Working Paper No. 33598 (Mar. 2025) (available at https://www.nber.org/system/files/working_papers/w33598/w33598.pdf) (last visited Jan. 22, 2026); *see also* Anthropic_0000597354, at 59 (copy of same produced by Anthropic).

content."[13] Jia and Nagaraj also state that "training data and content are the lifeblood of generative AI foundation models and are therefore very valuable in determining model performance," which Jia and Nagaraj assert is "evidenced by an *active licensing market* for such data."[14] Jia and Nagaraj even cite multiple real-world examples of content holders, like Reddit and academic publisher Taylor & Francis, who have licensed their content to Google for tens of millions of dollars, and they claim that their "work also holds implications for pricing models for training data[.]"[15]

34. Additionally, and as I will describe in more detail in a later section, Anthropic's own witnesses have testified that Anthropic does not value all types of data the same. *See also* Zhao Rebuttal Report ¶¶ 104, 114. For example, Anthropic considers [redacted] *See* below ¶¶ 71–72.

35. Basic economic principles also dictate that the incremental value of new and unique data should actually *increase* as an LLM company acquires more training data. Given Anthropic's own statements about the necessity for unique and diverse data to train its models, *see* Nagaraj Report ¶ 22, as Anthropic acquires data, its demand for unique data should commensurately increase, thereby raising the overall value of that data. That is, when Anthropic begins acquiring data to train a model, its data needs are necessarily less specific; once it has acquired a significant

[13] Anthropic_0000597354, at -59.

[14] *Id.* at -57.

[15] *Id.* at -56, -71.

amount of data, the value of data that is unique from the data it already has increases. Indeed, this is also what Dr. Nagaraj concluded in his *Cloze Encounters* paper.[16]

36. Further, even assuming that any individual piece of data, or category of data (like song lyrics), has minimal impact on *overall* performance of a model across all of its functions, that is not necessarily true with regard to specific use cases. To the extent Anthropic wants its Claude models to "be familiar with everything from how emails are drafted, how to write programs in C++, how many states are in the Union, and on and on,"[17] song lyrics likely have an outsized impact on the model's performance on specific use cases, like songwriting and lyrical interpretation—as further evidenced by the fact that companies developing AI models specifically for those use cases have attempted to license, and in some cases actually have licensed, Publishers' lyrics for precisely that purpose. *See* Smith Report ¶ 121; *see also* Zhao Rebuttal Report ¶¶ 102–15.

37. Last, Dr. Nagaraj's opinion that there is "very little incremental value" in any given piece of data is once again untethered from his stated assignment, which he says was to determine the "potential market for and value of" ***Publishers' lyrics***. Indeed, his opinion appears to be based only on the perspective of a specific type of buyer—an LLM developer amassing "an extremely large training corpus." Nagaraj Report ¶ 50. From this forced perspective, Dr. Nagaraj fails to account for the incremental value of Publishers' lyrics to other AI companies, including "single-purpose" and "niche" AI developers that may compete for the rights to this data.

---

[16] *Id.* at -59 (noting that findings of the paper "support more bargaining power for content providers who provide more unique content.").

[17] Nagaraj Report ¶ 20 (quoting Anthropic's Chief Science Officer Jared Kaplan).

38. Accordingly, Dr. Nagaraj's analysis of his "Market to Assemble All of the Data Needed to Train a General-Purpose LLM," does not offer reliable insight regarding Anthropic's effect on the potential market for and value of Publishers' copyrighted lyrics.

4. Dr. Nagaraj's opinion regarding prohibitive transaction costs is irrelevant and circular.

39. After inappropriately limiting his analysis to a "Market to Assemble All of the Data Needed to Train a General-Purpose LLM," Dr. Nagaraj next concludes that such a market cannot exist on account of "prohibitively high transaction costs." As I will explain, not only is this opinion, once again, irrelevant to Dr. Nagaraj's stated assignment of analyzing Anthropic's effect on the potential market for and value of Publishers' lyrics, but the opinion also collapses into circularity.

40. By narrowly defining his market to include as buyers only "companies like Anthropic that are building general-purpose LLMS," Dr. Nagaraj has, in effect, defined his market model to only include buyers that have chosen to attempt to obtain large quantities of copyrighted content without licensing it.[18] And because these buyers claim to require such vast amounts of copyrighted data to achieve their subjective goal of building a "capable" general-purpose LLM, Dr. Nagaraj concludes that the transaction costs for licensing all of this data would be prohibitively expensive—ergo, the market cannot exist or emerge, and therefore "Plaintiffs cannot have been economically harmed in such markets." Nagaraj Report ¶ 12. In other words, by limiting his analysis only to a small set of buyers who have all refused to participate in the relevant market, Dr. Nagaraj concludes that such a market could never exist. Such analysis is circular.

[18] Specifically, I refer to Anthropic (Claude), Google (Gemini), Meta (Llama), and OpenAI (ChatGPT), all of which, as I noted above, are currently being sued or have faced lawsuits for copyright infringement based on the same unlicensed AI training conduct.

41. Even if one accepts Dr. Nagaraj's assertion to the contrary, this circularity was not preordained. A general-purpose LLM developer, like Anthropic, could have decided to license copyrighted content first instead of obtaining it without licensing it.[19] 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.[20] The fact that some members of this artificially narrow buyer class have so far chosen not to do so does not prove that such a market cannot exist—it only proves that those firms have chosen not to enter that market.

42. Moreover, and once again, 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, and Dr. Nagaraj's opinion has no utility in determining the *value* of Publishers' lyrics. Dr. Nagaraj does not offer any evidence to support a conclusion that Anthropic, or any other buyer for that matter, has been unable to license Publishers' lyrics on account of prohibitively high transaction costs, and I am unaware of any such evidence in the record.

43. In fact, Tom Turvey, Anthropic's Vice President of Data Partnerships, testified that Anthropic has never approached Publishers to license any of their works, let alone enter into negotiations with them. T. Turvey 30(b)(6) Tr. 37:20–38:1. Mr. Turvey even testified that [redacted] T. Turvey 30(b)(6) Tr. 38:2–5, and that

---

[19] Indeed, this could even be done as a branding or market signifier, where a company commits to artists and content creators by licensing all of the copyrighted content in its training corpus, similar to how Anthropic purports to "distinguish[] itself from its competitors through its strong commitment to . . . ensuring that AI generates positive social impacts." *See* Nagaraj Report ¶ 15.

[20] Dr. Nagaraj even acknowledges that there is already an "active licensing market" for AI training data in his *Cloze Encounters* article. Anthropic_0000597354, at -57.

[redacted]

Anthropic_0000229152, at -52–53; T. Turvey 30(b)(6) Tr. 43:2 – 44:18. Accordingly, there is no evidence in the record that Anthropic attempted to license Publishers' works to be used as AI training data but was unable to do so on account of prohibitively high transaction costs.

44. Anthropic's policy of [redacted] also contradicts recent public comments from CEO Dario Amodei about the company's ethical responsibility to "think[] about the effects of the technology they build" and "not duck[] responsibility."[21] As I outlined in my initial report, there are many negative social and economic downstream effects that will likely arise from Anthropic's unmitigated ingestion and harvesting of copyrighted work, especially in the creative industries. *See, e.g.*, Smith Report ¶¶ 189–91. Dr. Nagaraj does not grapple with these negative effects in his report, nor the contradiction between Anthropic's policy of not paying for content it uses and its public comments about its ethical responsibilities.

45. By contrast, there is significant evidence that Publishers were able to overcome transaction costs associated with licensing their works to other AI companies, including Klay and Udio. *See* Smith Report ¶¶ 120–136. And Dr. Nagaraj himself claims in a co-authored paper that "a nascent market for licensing content" among LLM developers, including general-purpose LLMs, already exists, contradicting his claim here that such a market does not and cannot exist.[22]

---

[21] *Watch: Anthropic CEO Dario Amodei from World Economic Forum | WSJ*, YOUTUBE (Jan. 21, 2026) (available at https://www.youtube.com/watch?v=K7F6ohcBJus) (last visited Jan. 22, 2026); *see also* Keach Hagey, *Anthropic CEO Says Government Should Help Ensure AI's Economic Upside Is Shared*, WALL ST. J. (Jan. 20, 2026) (available at https://www.wsj.com/tech/ai/anthropic-ceo-says-government-should-help-ensure-ais-economic-upside-is-shared-1eab376c?st=38ZMHf).

[22] Anthropic_0000597354, at -62–63.

Indeed, in another paper, Dr. Nagaraj and his co-authors openly suggest that, given the unsettled state of copyright law, developers could "mitigate legal risks *by using data for which they have licenses* or explicit permission to use for training data"[23] (emphasis added). Dr. Nagaraj and his co-authors evidently agree that preventing lengthy and costly copyright litigation is sufficiently valuable to give rise to the emergence of a licensing market for AI training data in at least some circumstances.

46. There is also evidence that Anthropic's founders initially contemplated compensating rightsholders "for their labor, not at a flat rate, but with a fraction of the profits from the model produced."[24] In an internal Anthropic memo from July 2021, Dario Amodei brainstormed ways in which Anthropic could "value the contribution of data creators to AI models" and compensate them for that content, including suggesting that "data/content creators own equity in [the model] by virtue of their contribution to training it."[25] Indeed, it appears that Dr. Amodei originally envisioned compensation for rightsholders as "a sort of 'guaranteed basic capital' for the machine age."[26] These early contemplations show that Anthropic's founders recognized and understood that the value of copyrighted works as AI training data goes beyond mere token yield and should account for things like the cost of labor and the value of preventing

---

[23] Pierre Azoulay, Joshua L. Krieger, & Abhishek Nagaraj, *Old Moats for New Models: Openness, Control, and Competition in Generative AI*, NAT'L BUREAU OF ECON. R., Working Paper No. 32474, at p. 16 (Revised July 2024) (available at https://www.nber.org/system/files/working_papers/w32474/w32474.pdf) (last accessed Jan. 21, 2026).

[24] Anthropic_0000462964.

[25] Anthropic_0000462964, at -65.

[26] *Id.*; *see also* Dario Amodei 30(b)(1) Tr. 91:25–98:20 (testifying generally about his concern for "the concentration of wealth that comes from AI").

AI companies from becoming so "unpopular" that "[t]he government gets mad and cracks down on [them]."[27]

47. Beyond its circularity and contradictions, Dr. Nagaraj's conclusion regarding prohibitively high transaction costs in this market also relies on several flawed and irrelevant assumptions.

48. First, Dr. Nagaraj assumes that an LLM developer will not license any content if it cannot license all of the content "needed" to train a model. As discussed above, Dr. Nagaraj baked this assumption into how he defined the relevant market, but he uses it here to make the point that "[e]ven if Anthropic could execute 10 deals per month, with an average of 30 billion tokens per deal, it would take over [redacted] to negotiate enough deals to build a [redacted] corpus." Nagaraj Report ¶ 40. From this he concludes (without proof) that transaction costs are too high, and no market can exist. *See id.* But by Dr. Nagaraj's own admission, there are other *legal* ways to build a training corpus such that an LLM developer could feasibly enter the market to license specific types of data, including for "last-mile" and "gap-filling" data, and Dr. Nagaraj makes no attempt to discuss a situation where an LLM developer may decide to build a model with a smaller token corpus instead of the 40 trillion corpus he uses as an example.

49. Second, despite claiming that transaction costs prohibit the development of a market, Dr. Nagaraj ignores Anthropic's own role in increasing those transaction costs (and therefore ignores the solutions it has to reduce said transaction costs). As Mr. Turvey testified, [redacted] Anthropic_0000229152, at -55–56;

[27] Anthropic_0000462964, at -65.

T. Turvey 30(b)(1) Tr. 70:17–22. This lack of transparency *raises* transaction costs for rightsholders, who have to determine whether Anthropic is using their works for training, either through investigation, litigation discovery, or some other costly means. *See* Smith Report ¶¶ 154–56.

50. Third, Dr. Nagaraj ignores the burgeoning licensing market and the dozens of cases where other AI companies, including general-purpose LLM companies, have surmounted transaction costs. As I discussed in my initial report and described in the charts above (*see* ¶ 26), transaction costs clearly have not stopped other AI companies, including Anthropic's general-purpose LLM competitors, from entering into licensing agreements for AI training data. Anthropic itself has [redacted], suggesting that, when Anthropic is willing, transaction costs are not a barrier to licensing lyrics. *See* Anthropic_0000429496; *see also* T. Turvey 30(b)(6) Tr. 39:20–25. In his *Cloze Encounters* article, Dr. Nagaraj expressly acknowledges the development of "a nascent market for licensing content" and claims that his work "holds implications for pricing models for training data," which could help streamline licensing negotiations and reduce transaction costs.[28]

51. Dr. Nagaraj emphasizes the complexity of rights negotiations, *see* Nagaraj ¶¶ 45–46, but that is common and to be expected in new markets. Indeed, this was the case in the early days of essentially every new digital distribution channel, but in each case early entrants figured out how and what to negotiate, markets emerged, and future negotiations became more routine with lower transaction costs. As a result, today consumers enjoy licensed services for consuming music via online purchase, online streaming, Internet Radio, online video platforms like YouTube,

[28] Anthropic_0000597354, at -62, -71.

and for using licensed music via a variety of user-generated content platforms. Dr. Nagaraj does not account for any of this before jumping straight to his conclusion that a market cannot emerge in the future in this setting, nor does he explain why this setting is fundamentally different from prior licensing markets.

52. Dr. Nagaraj also ignores that the music publishing industry, in particular, has previously successfully developed solutions to overcome similar transaction-cost challenges. The music publishing industry is no stranger to this problem. More than a century ago, performing rights organizations evolved to intermediate licensing and royalty collection (and lower transaction costs) for the public performance of music compositions in bars, restaurants, arenas, and other public venues.[29] In the United States, performing rights organizations like the American Society of Composers, Authors and Publishers ("ASCAP") and Broadcast Music, Inc, ("BMI"), among others, collectively represent millions of rightsholders and administer licenses with hundreds of thousands of licensees.[30] Transaction costs in the public-performance context pose even greater challenges than in AI licensing because there are large numbers of parties on both the licensee and licensor sides of the market, whereas in the AI licensing market the potential licensees are much more limited. Yet those challenges have been overcome. And other collective rights organizations, like the Mechanical Licensing Collective, which administers a blanket mechanical license of musical compositions for streaming services, and SoundExchange, which collects and distributes

---

[29] WIKIPEDIA, *Performance rights organisation*, https://en.wikipedia.org/wiki/Performance_rights_organisation (last visited Jan. 23, 2026).

[30] ASCAP, *About ASCAP*, https://www.ascap.com/about-us (last visited Jan. 23, 2026); BMI, *About BMI*, https://www.bmi.com/about (last visited Jan. 23, 2026).

digital performance royalties for sound recordings to hundreds of thousands of recipients, have also surmounted similar challenges.[31]

53. ASCAP has suggested that a blanket license, "meaning that in exchange for a fee, the licensee may use any amount of music in the ASCAP repertoire . . . has long been known as the most efficient and cost-effective form of licensing at scale, and could be a reasonable form of licensing for AI."[32] The Copyright Clearance Center has also taken steps to surmount transaction costs through collective licensing.[33] I include these examples, not to endorse these proposals, but to highlight that industry participants in this market are actively considering and exploring potential solutions to the transaction-cost problem—solutions that Dr. Nagaraj wholly ignores.

54. Fourth, Dr. Nagaraj invokes irrelevant considerations that he says would increase transaction costs. For example, he claims that "the scale of training data required for general-purpose LLMs" means that "developers would need to make deals with hundreds of millions of different rightsholders." Nagaraj Report ¶ 42. But this is only true within Dr. Nagaraj's artificial and unsupported constraint that developers must license *all* the data used to train their model. In any case, licensing *Publishers'* copyrighted works would not require dealing with "hundreds of millions of different rightsholders." *Id.*

---

[31] THE MLC, *About The MLC*, https://www.themlc.com/our-story (last visited Jan. 23, 2026); SOUNDEXCHANGE, *About SoundExchange*, https://www.soundexchange.com/who-we-are/#about-us (last visited Jan. 23, 2026).

[32] ASCAP, *Comments Before the United States Copyright Office in the Matter of Artificial Intelligence and Copyright* (Oct. 30, 2023) (available at https://www.ascap.com/~/media/files/pdf/articles/2023/ascap-comments-to-us-copyright-office-on-generative-ai.pdf/) (last visited Jan. 22, 2026).

[33] RESEARCH INFORMATION, *CCC Launches AI Systems Training License* (Mar. 5, 2025) (available at https://www.researchinformation.info/news/ccc-launches-ai-systems-training-license/) (last visited Jan. 22, 2026).

55. Dr. Nagaraj next claims that the labor required by AI companies to locate and identify relevant counterparties and stakeholders "would be prohibitively expensive given the sheer number of rightsholders." *Id.* ¶ 43. But again, this is only true under Dr. Nagaraj's flawed assumption that AI models can only exist if they are trained with essentially all of the copyrighted content in existence, which as noted above is demonstrably false. *See id.* ¶¶ 40–42. In any event, this concern would not prevent a licensing market from emerging, as there are many easily identifiable rightsholders, including Publishers in this case, who hold the rights to a substantial market share of their respective industries. Indeed, Anthropic has already [redacted] *See* Smith Report ¶ 150 (discussing Anthropic_0000231226). Moreover, Anthropic could reduce the transaction costs associated with identifying the relevant rightsholders for works already contained in Anthropic's training corpus by making the contents of a potential training corpus fully transparent, as this would allow rightsholders to identify their content and initiate contact with Anthropic, easing the burden on Anthropic.

56. As for the "attribution" problem Dr. Nagaraj describes, Nagaraj Report ¶ 43, this is only a concern for AI companies that crawl, scrape, and download content from unlicensed sources. An AI company that licenses *before* training won't have this problem. In any case, attribution should not be an issue for identifying *Publishers'* works, as I understand they require their lyric-display licensees to include all relevant copyright infringement information ("CMI") alongside the lyrics when displayed online.[34] The CMI shows who the relevant counterparty for licensing would be, and a company like Anthropic that obtains the vast majority of its data from

[34] *See, e.g.*, K. Draughon 30(b)(6) Tr. 126:17–21 (testifying that Capitol CMG requires "lyric websites to which it licenses lyrics to include copyright management information . . . when they display lyrics").

the Internet can simply look at the copyright notices it has scraped. Expert Report of Shawn Shan (Dec. 19, 2025) (the "Shan Report"), at ¶¶ 57–59.

57. Dr. Nagaraj next claims that negotiations with identifiable counterparties may be "infeasible because a counterparty may possess only certain rights to a particular work." Nagaraj Report ¶ 44. Once more, this does not foreclose the existence of a market for copyrighted works: that some counterparties may lack the authority to license a work for AI training would not prevent those counterparties who possess that right from doing so. And again, Dr. Nagaraj fails to apply this point to *Publishers'* copyrighted works. Dr. Nagaraj's examples regarding books are also irrelevant to this matter, as are his second-hand descriptions [redacted] both of which have distinct rights-sharing frameworks as compared to music publishers, and which involve types of works not at issue here. *See id.* ¶¶ 45–46.

58. Dr. Nagaraj's assertion that any particular rightsholder has only "a fraction of the data needed to train a general-purpose LLM such as Claude," is also irrelevant. *Id.* ¶ 47 Dr. Nagaraj does not explain why any particular rightsholder would need to be able to license more than a fraction of the needed data for a training-data market to exist. While the inability of any single rightsholder to license all or the majority of the data needed to train a model like Claude means that a general-purpose LLM developer would need to license data from multiple rightsholders, Dr. Nagaraj has not shown such a state of affairs multiplies transaction costs beyond the feasible. In any event, what a specific company subjectively needs to meet a specific business objective does not foreclose the existence of a market for copyrighted works, and certainly not for Publishers' lyrics, as explained above. Dr. Nagaraj fails to explain why a company like Anthropic could not simply attempt to make a model with less than the ideal amount of copyrighted data, or

why it could not rely on public domain or even synthetic data for training, while supplementing the remainder through so-called "last-mile" and "gap-filling" deals.

59. Dr. Nagaraj also claims that "incredibly time-consuming" rights negotiations "will not be feasible" at an immense scale. Nagaraj Report ¶ 48. Again, this conclusion rests on the flawed assumption that an LLM developer will only attempt to build a training corpus if it can license from "hundreds of millions of rightsholders." *Id.* ¶¶ 42, 48. Dr. Nagaraj cites no evidence that negotiations with *Publishers* would be "incredibly time-consuming," including because Anthropic admitted that it has never even attempted to reach out to *any* music publishers for licensing, let alone Publishers in this case. T. Turvey 30(b)(6) Tr. 37:20–38:5. Despite citing select evidence where Anthropic's Mr. Turvey appears to have been left waiting at various stages of negotiation with different counterparties, Dr. Nagaraj ignores the more than [redacted]. *See* Anthropic_0000441006 [redacted] *see also* T. Turvey 30(b)(6) Tr. 43:19–21 (stating that he began working at Anthropic on February 20, 2024). And finally, Dr. Nagaraj does not discuss how rights negotiations become increasingly simpler and more routine as a market develops and a new technology becomes better understood, as with negotiations with digital distribution platforms like Spotify, YouTube, TikTok, and other platforms that now regularly and routinely renew licensing contracts.

60. Last, and as I will address in more detail below, Dr. Nagaraj ends his transaction-cost discussion by claiming that "these significant transaction costs far exceed the value that any individual piece of training data could have," and therefore that "no actual or potential market for data required in a general-purpose LLM's training corpus is likely to be viable in the face of these transaction costs." Nagaraj Report ¶¶ 50–51. Tellingly, however, Dr. Nagaraj does not conclude

from here that, in light of these "prohibitive" transaction costs, Anthropic should get out of the LLM-development business altogether, nor does he opine that the market for general-purpose LLMs is inherently unsustainable, and the technology should be abandoned. Instead, he implicitly concludes that rightsholders, including Publishers here, should simply relinquish their works to Anthropic *for free*. In other words, because he believes it would be too costly for Anthropic and other LLM developers to license all the material they need to build their product, Dr. Nagaraj concludes that developers should not have to pay *anyone anything* to use copyrighted works as training data.

61. Dr. Nagaraj makes no attempt to grapple with the immense social and economic consequences that would flow from this conclusion. For example, he does not discuss how rightsholders, including Publishers, would be incentivized to pull down their work from the Internet and impose stringent paywalls on websites to prevent AI companies—and, as collateral damage, consumers—from accessing content without payment. Taken to its logical conclusion, Dr. Nagaraj's opinion supports a framework where consumers pay the price for Anthropic's AI training, while Anthropic reaps the economic benefit from the creative industry's cost and labor used to create the copyrighted works that its models consume.

5. Dr. Nagaraj appears to further limit his analysis to the year 2021.

62. At points in his report, Dr. Nagaraj appears to cabin his analysis to whether a licensing market for AI training data existed in 2021. *See, e.g.*, Report ¶ 53 ("The low value of lyrics as training data, paired with the substantial transaction costs of negotiating licenses, lead me to conclude that a market for lyrics training datasets did not exist in 2021 and is not likely to form in the future."). To the extent he does so, this is a significant error.

63. There are obvious problems with limiting analysis to only 2021. Based on the Amended Complaint (ECF No. 337), I understand that the alleged infringements in this case took

place as late as 2025, which means that Dr. Nagaraj's opinions are irrelevant to much of the alleged conduct at issue in the case. Similarly, most of the evidence that he relies on to form his opinions is from after 2021 (for example, Mr. Turvey was only hired in February 2024, and yet Dr. Nagaraj relies on much of his testimony to justify why a market did not exist in 2021). And perhaps most importantly, the technology at issue here is so new that Anthropic was not even founded until 2021, *see* Nagaraj Report ¶ 15, and the first publicly available LLM product (OpenAI's ChatGPT) was not released until November 2022.[35] To expect a fully-formed licensing market to develop within months of a new technology's public introduction public is unrealistic—and certainly does not justify concluding that such a market "is not likely to form in the future." *Cf.* Smith Report ¶¶ 110–113 (describing how the AI-licensing market parallels other examples of new technology in modern history).

64. By cabining his analysis to 2021, Dr. Nagaraj also ignores all the ways in which the licensing market for AI training data has already developed. Indeed, the earliest training data licensing listed in Exhibit A to my initial report is from October 2022. To the extent Dr. Nagaraj intended to cabin his opinions to the state of the market in 2021, he ignores a wealth of relevant evidence contradicting his views and supporting the case that—as he himself observed in his *Cloze Encounters* paper—an active licensing market for such data exists and is thriving.[36]

B. Dr. Nagaraj's "Factor Four" Analysis of a Market for Lyrics Used to Train a General-Purpose LLM is Flawed.

65. The second purported market that Dr. Nagaraj analyzes is just as flawed as the first. Specifically, he attempts to provide a "factor four" analysis of a "market for ***only*** lyrics data to

---

[35] WIKIPEDIA, *ChatGPT* (https://en.wikipedia.org/wiki/ChatGPT) (last visited Jan. 23, 2026).

[36] Anthropic_0000597354, at -57.

train a general-purpose LLM." Nagaraj Report ¶ 52 (emphasis added). Once again, by its very terms, this market is divorced from Dr. Nagaraj's stated assignment of analyzing "the effect of [Anthropic's] use upon the potential market for and value of [Publishers'] copyrighted work[s]." *Id.* ¶ 6. And once again, the constraints and flawed assumptions underlying Dr. Nagaraj's use of this market infects and skews his subsequent "factor four" analysis.

1. Dr. Nagaraj uses the same constrained and unreliable market parameters regarding General-Purpose LLM and buyer "needs."

66. At the outset, Dr. Nagaraj's alternative market suffers from many of the same underlying flawed assumptions that skewed his analysis of the first market. *See* above § II.A. For example, once again, Dr. Nagaraj artificially constrains his market to only include "buyers" like Anthropic building "general-purpose LLMs,"—notwithstanding that the "potential market for and value of" Publishers' lyrics are determined by a much broader range of potential buyers. *See* above II.A.2. As discussed above, it makes no economic or logical sense to define the relevant market this way.

67. Similarly, Dr. Nagaraj continues to frame his market in terms of what an LLM developer "needs" to accomplish its goal. *See, e.g.*, Nagaraj Report ¶ 52 ("A market for only lyrics data to train a general-purpose LLM simply could not form because the developer would not be able to train its model without the massive quantities and varieties of data it ***needs***.") (emphasis added). For the same reasons I described above, this is a flawed assumption that skews Dr. Nagaraj's subsequent "factor four" analysis and leads him to numerous misleading conclusions that are irrelevant to or are unreliable in view of his own stated assignment. *See* II.A.3.

2. Dr. Nagaraj unnecessarily frames his alternative market in terms of General-Purpose LLM developers attempting to train their models with *only* song lyrics.

68. For Dr. Nagaraj's second market, he unnecessarily frames his analysis in terms of general-purpose LLM developers who use "***only*** lyrics data to train a general-purpose LLM." Nagaraj Report ¶ 52 (emphasis added). Specifically, he concludes that such a market "simply could not form because the developer would not be able to train its model without massive quantities and varieties of data it needs." *Id.* (emphasis added). This is a strange inclusion; whether a general-purpose LLM could be trained only on song lyrics is not at issue in this litigation, so Dr. Nagaraj's conclusion is irrelevant.

69. Once again, the correct market that Dr. Nagaraj should be analyzing is the market for ***Publishers' copyrighted works***. His failure to properly frame the relevant market renders all of his subsequent analysis in this section misleading and unreliable in determining what effect Anthropic has had on the "potential market for and value of" Publishers' copyrighted works.

3. Dr. Nagaraj's opinions regarding the valuation of music lyrics rest on erroneous and conflicting assumptions.

70. First, Dr. Nagaraj erroneously assumes that music lyrics are low in value because they yield a relatively low number of tokens. *See* Nagaraj Report ¶¶ 55–56. But this way of valuing lyrics only makes sense if one assumes that all training data is valued equally. In the very same section of his report, however, Dr. Nagaraj admits that some training data is more valuable than other data. *Id.* ¶ 58. And in his *Cloze Encounters* paper, he "refute[s] the claim that any single source has a negligible effect on model performance, even in large production models."[37] As I described in my initial report, many of Anthropic's own employees have testified to the value and

[37] Anthropic_0000597354, at -59.

utility of song lyrics as training data and have used Claude for various song- and lyric-related uses, including, for example, to generate "new" song lyrics. *See* Smith Report ¶¶ 81–83.

71. Dr. Nagaraj also admits that "high-quality textual data is essential for training an LLM." Nagaraj Report ¶ 22. Indeed, this is why Anthropic values [redacted] T. Turvey 30(b)(6) Tr. 25:19–22 [redacted] *see also* Shan Report ¶ 40.iv (finding that [redacted] and that [redacted] This is also why Anthropic "cleans" its data to remove less-valuable data that may impair or negatively affect an LLM's performance. *See* D. Drain 30(b)(1) Tr. 117:25–119:24 (discussing importance of removing "garbage" data from Common Crawl dataset during preprocessing phase of model training).

72. Dr. Nagaraj's reliance on Dr. Bilmes's belief that "lyrics are themselves not a diverse dataset for purposes of training a general-purpose LLM" is also starkly contradicted by Anthropic's Chief Science Officer, Jared Kaplan. Specifically, Dr. Kaplan testified in this case that lyrics are useful to LLM training because they provide metaphor, understanding of language, artistic expression, and nuance.[38] To that end, Dr. Nagaraj states in his report that the aim of LLM training is to "captur[e] the variety of linguistic patterns, interactions, and nuances of human language," Nagaraj Report ¶ 22, all of which song lyrics can and do express. *See* Zhao Rebuttal Report ¶¶ 102–15.

[38] *See* Smith Report ¶ 56 (quoting Aug. 22, 2024 Declaration of Dr. Jared Kaplan (ECF No. 209 ¶ 35)).

73. Accordingly, it also makes no sense to judge the value of music lyrics as AI training data by their utility in responding "to generalized questions." *Id.* ¶ 57. Their value is primarily rooted in artistic expression, linguistic nuance, and cultural importance. Indeed, Anthropic's employee, Dawn Drain, testified that it was Anthropic's intent for Claude to fully memorize certain "pillars of human culture," which she agreed could include such iconic songs like "Let it Be" by The Beatles, or "American Pie" by Don Mclean (a work at issue in this litigation). D. Drain 30(b)(1) Tr. 107:24–109:11.

74. Second, Dr. Nagaraj's valuation of music lyrics is infected by an underlying assumption that there is a free and unlimited supply of music lyrics that can be used in AI training. Specifically, to support his conclusion that music lyrics have relatively low value as AI training data, Dr. Nagaraj relies heavily on an internal spreadsheet reflecting the relative valuations of training data from three members of Anthropic's technical staff. Nagaraj Report ¶¶ 58–59. The spreadsheet contains [redacted] Anthropic_0000228080.

75. At the outset, it is notable that Anthropic included [redacted]

[39] Dr. Nagaraj appears to contend that the "Music" category, as it is listed in the spreadsheet, relates to the music *lyrics* at issue in this case, however, he notes some uncertainty about that in a footnote. *See* Nagaraj Report ¶ 59 n.93. For purposes of this rebuttal, I assume that the "Music" category includes lyrics, but I reserve the right to revise that assumption if new facts come to light.

[redacted] which at most represent anecdotal evidence about Anthropic's valuation of lyrics, not the wider market's valuation of lyrics (which is what Dr. Nagaraj should be analyzing).

76. More fundamentally, the spreadsheet presumably reflects Anthropic's [redacted] Because Anthropic's employees are presumably [redacted].[40] If the available "supply" of music lyrics suddenly fell to zero, and Anthropic could no longer scrape any music lyrics from the Internet, economic principle dictates that [redacted] This also comports with Dr. Nagaraj's findings in his *Cloze Encounters* paper that rare, unpopular, and unique works are actually *more* valuable in LLM training given their lack of accessibility (*i.e.*, there is a lower available supply), which Dr. Nagaraj concludes "support[s] more bargaining power for content providers who provide more unique content."[41] Indeed, when boiled down to its essence, the logical argument being presented by Dr. Nagaraj here is: Anthropic can get copyrighted lyrics for free on the internet without a license, Anthropic doesn't pay for things it can get for free, therefore

---

[40] Indeed, even Dr. Nagaraj acknowledges in his report that this spreadsheet was [redacted] Nagaraj Report ¶ 58 (emphasis added).

[41] Anthropic_0000597354, at -59.

Anthropic isn't part of the market, therefore Anthropic's use of the unlicensed copyrighted lyrics can't harm the licensing market for lyrics. But in the context of this case, that framing is a classic example of begging the question: it assumes that obtaining and using unlicensed copyrighted materials for free is legitimate, when in fact that legitimacy is a central question at issue in this case.

77. Dr. Nagaraj's conclusion that "there are other ways to help train LLMs to answer user prompts regarding songs and songwriting, " Nagaraj Report ¶ 60, is insufficiently supported. His only specific support for this assertion appears to be quotes from various Anthropic deponents that songwriting can be learned from "music reviews" and "books about how people write songs." *Id.* Dr. Nagaraj does not cite any actual technical analysis of the operation of LLMs that shows that an LLM not trained on lyrics could respond to the same prompts regarding songs and songwriting at the same quality that an LLM trained on lyrics can answer. *See also* Zhao Rebuttal Report ¶¶ 109–10. In any case, either Anthropic does not need lyrics to train its model—in which case it has no need to scrape and use them for that purpose—or it does, in which case the lyrics should be sufficiently valuable to justify licensing.

78. Third, Dr. Nagaraj's opinion regarding "lyric-related research" and the frequency with which Claude users ask about songs and songwriting is, at most, relevant to Anthropic's internal valuation only, not the market valuation of Publishers' lyrics specifically. Nagaraj Report ¶ 61. Moreover, the evidence that Dr. Nagaraj relies on to make this point is stated in relative terms among total Claude usage, not in absolute terms of total uses. For example, Dr. Nagaraj claims that an internal Anthropic "Clio" report "confirms that none of the top ten use cases for Claude.ai, which account for more than 60% of the use cases analyzed, involves requesting lyrics." *Id.* But this observation says nothing about how many times per month Claude users look up lyrics, which,

given the scale of queries involved in this case alone, could be several hundred thousand and still not crack the "top ten" of Claude use cases. *See, e.g.*, Smith Report ¶¶ 88–91 (reflecting expert Josh Dennis's estimations that over a six-month period there were 200,000 song-related Claude queries). To that end, Dr. Nagaraj's inclusion of his own personal usage of AI is baffling—how he personally uses AI is not reliably probative of anything meaningful about AI use trends.

79. Regardless of all the foregoing, Anthropic's evidence about its own valuation of lyrics represents only one market participant's valuation. Dr. Nagaraj's stated assignment requires analysis of valuations from all buyers of Publishers' lyrics as well as Publishers' valuation of their lyrics. Even Anthropic's own CEO and chief founder, Dario Amodei, recognized that a market for AI licensing data must account for both the buyer's valuation *and* the seller's valuation.

80. Specifically (and as discussed during Dr. Amodei's January 8, 2026 deposition), in July 2021, Dr. Amodei drafted an internal Anthropic memo titled "An Economic Model for Compensating Data Producers."[42]In the document, Dr. Amodei brainstorms how to "value the contribution of data creators to AI models," which he proposes doing by, among other things, "calculate[ing] the marginal impact of data vs compute on the performance of the model."[43] Another Anthropic employee, Andy Jones, however, correctly points out in a comment that this calculation "gives a plausible mechanism for [Anthropic] to come up at a price at which we'd *buy* data, but to make a market it needs to be paired with a mechanism for figuring out what price people'd *sell* their data at."[44] Dr. Amodei agrees: "Having only one side of the market is like forcefully buying the data by eminent domain. I suppose the price for selling comes from the

---

[42] Anthropic_0000462964; *see* Dario Amodei 30(b)(1) Tr. 89:10–22.

[43] Anthropic_0000462964.

[44] *Id.* at -97

implicit threat content providers have to disallow training on their data—not sure how we would calculate that, but seems worth doing something like it."[45]

81. In other words, Anthropic's own CEO acknowledges that a market for training data cannot be determined solely by the buyer's willingness to pay for data and must also account for the price at which the seller will actually sell. I describe the same concept in my initial report as a necessary precondition for any prospective licensing market. *See* Smith Report ¶ 107 (describing "bargaining set" and need for overlap in valuation between buyers and seller). Dr. Nagaraj, however, does not discuss the "seller's" valuation of data anywhere in his report, ostensibly because he believes there is *no* price at which a "buying" general-purpose LLM company would be willing to pay for training data, making the seller's valuation irrelevant. But this is wrong because it erroneously assumes that the buyer already has access to an available "free" source of data through its scraping and crawling practices.

4. Dr. Nagaraj's opinions about transaction costs for licensing music lyrics are speculative and irrelevant.

82. Dr. Nagaraj's conclusions regarding transaction costs in his alternative market are also flawed and untethered from his stated assignment of analyzing the market and value of *Publishers'* copyrighted lyrics. Specifically, his conclusion that the transaction costs for licensing music lyrics "would be prohibitively expensive" is speculative, and he makes no attempt to analyze transaction costs associated with licensing *Publishers'* lyrics. Nagaraj Report ¶ 65.

83. Dr. Nagaraj begins by stating that there are "over a hundred million songs protected by copyright" and that identifying each rightsholder would be "extremely burdensome, if not impossible." *Id.* ¶ 66. I do not understand the relevance of this observation to the potential market

[45] *Id.*

for or value of Publishers' copyrighted works. Dr. Nagaraj speculates that licensing one hundred million songs would involve a "long tail" of musicians from which Anthropic would need to acquire rights, *see id.*, but he bases this on generalities of the music industry, not on any fact in evidence; he does not cite any evidence that Anthropic actually trained any of its models on lyrics from this "long tail" of songs, such that Anthropic would need to license those works. In any case, I understand from counsel that Publishers here own all the rights necessary to license their lyrics as AI training data, so this concern is irrelevant to analyzing the potential market for and value of Publishers' lyrics.

84. Dr. Nagaraj next opines that valuing information is hard, and agreements related to new technologies are complicated and take time to negotiate, all of which supposedly increases transaction costs. *Id.* ¶¶ 67–70. Dr. Nagaraj ignores that Anthropic did not even attempt to engage in licensing negotiations with Publishers, *see* T. Turvey 30(b)(6) Tr. 37:20–38:1, which makes it impossible to know how any such negotiations would have played out in reality. Dr. Nagaraj's opinion that those negotiations would have been difficult is pure speculation. While Dr. Nagaraj cites Publishers supposed "difficulties" in negotiations with *other* AI companies,[46] he once again ignores the fact that Publishers ultimately did succeed in negotiating and licensing training rights to other LLM companies, which means they surmounted any challenges associated with transaction costs. As is typical with new technologies, and as noted above, Publishers' negotiators will likely learn from these initial demonstrated successes and be able to streamline future negotiations as the market matures. Dr. Nagaraj does not account for these likely developments.

---

[46] *See* Nagaraj Report ¶ 69. While Dr. Nagaraj claims in this paragraph that UMPG "struggle[d] with the basic terms of deals for training datasets," he omits that UMPG and Udio actually did come to an agreement. *See* Smith Report ¶ 121.

85. As mentioned above, Dr. Nagaraj again does not account for ways in which Anthropic increases transaction costs, nor the ways in which Anthropic could reduce transaction costs. *See* above ¶ 49. For example, Anthropic has a [redacted] *See* T. Turvey 30(b)(1) Tr. 70:17–22. I understand from counsel that Publishers here were only able to confirm that Anthropic did in fact use their works to train Claude through legal investigations and discovery in this case.

86. Accordingly, I believe Dr. Nagaraj's analysis of an alternative market is deeply flawed and leads to irrelevant and misleading conclusions regarding the valuation of and transaction costs associated with licensing Publishers' lyrics. Dr. Nagaraj's opinions about whether transaction costs make licensing orphan works impossible is also irrelevant.

C. Dr. Nagaraj's Opinions Regarding the Relevance of Existing Deals to License AI Training Data Contradict His Market Opinions.

87. After concluding that "no actual or potential markets existed or were likely to develop for either the data required to train a general-purpose LLM or lyrics data used to train a general-purpose LLM," Dr. Nagaraj attempts to explain why existing licensing deals do not invalidate his prior opinions. Nagaraj Report ¶¶ 72–81. The opinions in this section, and the mere existence of the deals Dr. Nagaraj discusses therein, contradict his previous opinions in his report.

88. First, Dr. Nagaraj expressly admits that "companies that develop LLMs could enter into deals to acquire data" under certain circumstances. *Id.* ¶ 72. This alone confirms that a market for AI training data exists and comports with Dr. Nagaraj's prior published observations in his *Cloze Encounters* paper about "the birth of a nascent market for licensing content."[47] Dr. Nagaraj

[47] Anthropic_0000597354, at -62.

even notes that Anthropic itself has [redacted] Nagaraj Report ¶ 73.

89. Dr. Nagaraj's conclusion that "such deals can only emerge in the context of a training dataset for a general-purpose LLM once developers have *already* assembled large enough datasets for its training corpus" is speculative and unsupported by evidence. Nagaraj Report ¶ 72 (emphasis in original). Dr. Nagaraj offers no compelling justification for why an LLM company would not or could not attempt to begin its training corpus with licensed content *and then* supplement with synthetic data or public domain data, let alone the possibility that a company might actually commit to only using licensed content. In any event, Dr. Nagaraj does not explain why a license for Publishers' works could not emerge in the context he identifies.

90. Relevant to a market for Publishers' lyrics, specifically, Dr. Nagaraj expressly acknowledges the "existence of certain agreements to license music for training data in certain LLMs," although he fails to specifically cite to or name them. Nagaraj Report ¶ 74. He also claims to be aware of other LLMs that have licensed text data from numerous entities across multiple industries, including book publishers, news organizations, and even "music data from music companies." *Id.* ¶ 75. While he claims to be unaware of any AI agreements that "involve licensing of ***lyric*** data to train general-purpose LLMs," *id.* ¶ 79 (emphasis in original), he ignores the fact that AI licensing deals for lyric data do still exist (and are therefore relevant to Dr. Nagaraj's stated assignment), including UMPG's deals with Udio and Klay, both of which expressly grant the right to use lyrics for training purposes.

91. Dr. Nagaraj attempts to discount other publicly-reported deals with developers of general-purpose LLMs that supposedly involve "non-training" use cases based solely on the contents of corporate press releases. *Id.* ¶ 76. He claims that other deals, like the licensing

agreement between Google and Reddit, "encompass a much larger scope of services beyond simply training general-purpose LLMs," and therefore "do not support the theory that a market for data sufficient to train a general-purpose LLM could form." *Id.* ¶ 76. Dr. Nagaraj ignores that a deal that includes a license for training in addition to other rights is still evidence of an existing licensing market. In fact, Dr. Nagaraj cited the Google-Reddit deal in his *Cloze Encounters* article as evidence for determining "the economic value of access to certain bits of training data."[48]

92. Dr. Nagaraj's attempts to distinguish licensing deals made for the benefit of general-purpose LLMs and deals made for "other types of generative AI companies that offer niche, multi-media AI products" are also unpersuasive. Nagaraj Report ¶ 77. Dr. Nagaraj does not explain why the different end-user services that each model offers mean the models are not competing for the same training data in the same market. On the contrary, a recent paper co-authored by Dr. Nagaraj asserts that foundation AI models like "Google's Gemini, Large Language Model Meta AI (Llama) by Meta, and Claude from Anthropic" serve as "the substrate developers use to build a wide range of generative AI applications," and "[g]iven their central position in the technology stack . . . can be a chokepoint for innovation and competition at the application level[.]"[49] Relevant here, Dr. Nagaraj and his co-authors note that this "application level" includes everything "from the mundane (automating email responses and shopping lists), to artistic (writing poetry, bedtime stories and screenplays), and skill building (e.g., personal tutors, coding

---

[48] Anthropic_0000597354, at -62–63.

[49] Pierre Azoulay, Joshua L. Krieger, & Abhishek Nagaraj, *Old Moats for New Models: Openness, Control, and Competition in Generative AI*, NAT'L BUREAU OF ECON. R., Working Paper No. 32474, at p. 5, 8 (Rev. July 2024) (available at https://www.nber.org/system/files/working_papers/w32474/w32474.pdf) (last visited Jan. 21, 2026).

copilots)."[50] On this economic and technological backdrop, it is hard to see how Dr. Nagaraj could conclude that general-purpose LLMs and single-purpose LLMs are *not* competitors in the same market for AI training data.

93. Similarly, Dr. Nagaraj does not explain the significance of "different deal structures or pricing" when it comes to "single-purpose AI companies" as compared to "general-purpose LLM" companies. Nagaraj Report ¶ 77. To the contrary, because they compete in the same market for training data, it is an error to acknowledge that a licensing market exists for "single-purpose" or "niche" AI models but not for "general-purpose LLMs." The use cases for general-purpose models necessarily encompass the use cases for "single-purpose" models and the general-purpose and single-purpose models compete with each other for customers drawn to those uses; it follows that these companies compete with each other for the data necessary to train models for those uses as well. Importantly, from the seller's perspective, there is no reason to differentiate between the two.

D. Dr. Nagaraj Ignores the Effect of Anthropic's Use on Other Relevant Markets.

94. As I have repeatedly emphasized, to adequately "analyze economic issues related to . . . 'the effect of the use upon the potential market for or value of the copyrighted work,'" Dr. Nagaraj should have framed his market in terms of the actual and potential market for ***Publishers' lyrics***. *See* Nagaraj Report ¶ 6 (stating assignment for his report). By instead framing his principal and alternative markets in terms of "All of the Data Needed to Assemble a General-Purpose LLM" and a "market for only lyrics data to train a general-purpose LLM," Dr. Nagaraj ignores numerous important economic factors and other relevant markets affected by Anthropic's alleged infringement.

---

[50] *Id.* at p. 1.

95. First and foremost, Dr. Nagaraj wholly ignores the potential existential threat that general-purpose LLMs like Anthropic's Claude poses to human artists and songwriters. *See* Smith Report ¶¶ 60–98. In his *Cloze Encounters* paper, Dr. Nagaraj expressly acknowledged that the emergence of LLMs presents "critical economic questions surrounding substitution and the potential for LLM outputs to cannibalize the market for original source material,"[51] and yet he does not offer so much as a sentence to that potential harm in his report. Instead, he appears to claim that, because a company like Anthropic cannot license all of the copyrighted works that it would ideally use to train its Claude model, a market for those works as training data does not and cannot exist, and therefore "Plaintiffs cannot have been economically harmed in such markets." Nagaraj Report ¶ 12. I do not understand how Dr. Nagaraj can acknowledge the danger of AI cannibalization in his academic writings yet claim in his expert report that Publishers will not suffer harm if Anthropic is able to train on their copyrighted works without license.

96. Second, in my initial report I discussed how Anthropic's unlicensed content-acquisition strategies and subsequent use of Publishers' lyrics prejudicially impacts Publishers' negotiating power in several different markets, including negotiations with other "single-purpose" AI companies like Udio and Klay. *See* Smith Report ¶ 159. Despite citing to some of these AI companies and acknowledging the existence of these types of licensing agreements, Dr. Nagaraj makes no attempt to analyze how Anthropic's unlicensed content acquisition and subsequent use of Publishers' lyrics could affect similar or future negotiations with other AI companies. Crucially, Dr. Nagaraj also fails to address how Anthropic's unlicensed content-acquisition practices would influence the incentives of other AI companies (of all varieties) to copy that conduct in order to

[51] Anthropic_0000597354, at -71.

compete with Anthropic. *See* Smith Report ¶ 159. To that end, Dr. Nagaraj also fails to address how Anthropic's conduct affects the related market for AI-generated derivatives.

97. Third, Dr. Nagaraj does not address how Anthropic's conduct undermines Publishers' ability to restrict access to their own content. I discuss the importance of this necessary market precondition at length in my initial report, and I support that discussion with citation to the U.S. Copyright Office's roundtable economic report "Identifying the Economic Implications of Artificial Intelligence for Copyright Policy" to which Dr. Nagaraj is a co-author. *See* Smith Report ¶¶ 102–06. Despite his discussion of transaction costs, Dr. Nagaraj offers no discussion about how Anthropic's mass crawling and scraping practice eviscerates Publishers' practical ability to restrict access to its content, and he offers no viable alternative technological or practical solution for rightsholders who wish to restrict access to their content by preventing Anthropic's mass scraping conduct. *See id.* ¶¶ 142–45.

98. To that end, Dr. Nagaraj does not even attempt to grapple with the negative social and economic implications that would follow if Anthropic's conduct becomes widespread. If Publishers cannot restrict access to their content via copyright enforcement, rightsholders of all kinds will have strong incentives to pull their works off the Internet to prevent scraping. Similarly, licensed content aggregators may impose paywalls or other viewing restrictions to content that was previously free to view online in order to prevent companies like Anthropic from taking content for free. And, as I discussed in my initial report, there would also likely be a substantial reduction in the intrinsic incentive among artists and creators to make new content if it can be taken for free and fed to a machine that may ultimately replace them. *See id.* ¶¶ 189–91.

99. Beyond that, Dr. Nagaraj also does not grapple with any of the potential negative implications posed by generative AI to the labor market, including Publishers' market for artists

and songwriters. As I discuss in my initial report, many of Anthropic's founders have predicted that generative AI will have devasting effects on the jobs and human labor markets. *See id.* ¶ 79. In his recent deposition in this case, Anthropic CEO Dario Amodei reaffirmed his belief that AI technology will concentrate wealth in AI companies and soon "make human labor obsolete," including the human labor of songwriters, because the technology will "eventually automate everything." Dario Amodei 30(b)(1) Tr. 98:4–20. Even more recently, Dr. Amodei spoke at the World Economic Forum in Davos, where he admitted that "the signature of this technology is it's gonna take us to a world where we have a very high GDP growth and potentially also very high unemployment and inequality."[52] Dr. Nagaraj does not address any of these externalities, nor how they will likely affect Publishers' business, and yet he categorically concludes that Publishers' will not suffer any harm as a result of Anthropic's alleged use and infringement of their copyrighted works to train its models.

E. Dr. Nagaraj's Regression Analyses Contain Numerous Flaws and Obscure the Relationship Between Claude and Lyric Interest.

100. Dr. Nagaraj performs regression analyses of two different data sources, website traffic and Google Trends data, which he adduces as support for his claim that Claude.ai did not harm Plaintiffs in the existing market for the online display of lyrics. Nagaraj Report ¶ 83. Specifically, Dr. Nagaraj asserts that there was no evidence of a "relationship between the introduction of Claude and a decrease in traffic to lyrics websites" or that "increased usage of Claude [caused] a decrease in lyric-specific Google searches." *Id.* ¶ 84.

---

[52] Ravikash Bakolia, *Anthropic CEO Talks Potential IPO, Gemini and AI Tech at Davos*, SEEKING ALPHA (Jan. 20, 2026) (available at https://seekingalpha.com/news/4540510-anthropic-ceo-talks-potential-ipo-gemini-and-ai-tech-at-davos) (last visited Jan. 22, 2026).

101. Dr. Nagaraj claims that under Plaintiffs' theory of harm, he would expect increased usage of Claude "to cause a decrease in traffic to [websites that display lyrics]" and "to cause a decrease in lyric-specific Google searches." *Id.* (emphasis added). Dr. Nagaraj claims that he "do[es] not find evidence" of either relation based on the results of his regression analyses. *Id.*

102. Based on my review of Dr. Nagaraj's methodology and models, I find that Dr. Nagaraj's conclusion that "Anthropic's alleged conduct did not harm the existing market for online lyrics display" is not based on reliable scientific principles and methods. *Id.* ¶ 148. As I detail in subsequent sections, the purported absence of evidence of either relation in Dr. Nagaraj's regression models could instead be a consequence of improper manipulation of data, unrealistic and unsubstantiated assumptions, and inappropriate model specification.

1. As designed, Dr. Nagaraj's regression models are incapable of evaluating the causality that he purports to analyze.

103. As discussed herein, neither Dr. Nagaraj's web traffic nor his Google Trends regression analyses are able to establish the causal connection that he claims to exist.

104. The fundamental methodological flaw in Dr. Nagaraj's analyses is most apparent in his regression models purporting to evaluate whether increased Claude usage "cause[d] a decrease in traffic to [websites that display lyrics]." Nagaraj Report ¶ 84. Dr. Nagaraj asserts that a statistically insignificant coefficient for his Claude.ai variable indicates that Claude did not have a significant effect on online lyrics consumption. *Id.* ¶ 121.

105. Dr. Nagaraj appears to be attempting to measure whether or to what extent Claude.ai serves as a substitute for the lyric websites he includes in his analysis, such that some consumers use Claude.ai to view lyrics, rather than visiting lyric websites. However, as designed, Dr. Nagaraj's model is not capable of establishing whether Claude is a substitute for lyric websites.

106. In fact, even under the assumption that consumers see Claude as perfectly interchangeable to lyric websites, such that every additional visit to Claude.ai corresponds with one fewer visit to a lyric website, Dr. Nagaraj's model could potentially produce an insignificant or significant, positive coefficient for his Claude.ai variable. This is a result of Dr. Nagaraj's decision to regress a measure of quantity on another measure of quantity, specifically regressing lyric website traffic on Claude.ai traffic. The issue with this approach is that even for substitute goods, a positive demand shock (*i.e.*, an increase in demand) could be expected to increase the quantity consumed for both goods.

107. In this context, a hypothetical song might "go viral" (a positive demand shock) and cause one consumer to visit a lyric website and another consumer to visit Claude, increasing traffic to both. Such a relation would be associated with a positive coefficient in his lyric traffic model and Dr. Nagaraj would conclude that there is no causal connection between Claude and lyric websites. Yet, in such a hypothetical scenario, potentially both consumers would have visited the lyric website absent Claude being an available alternative.

108. Dr. Nagaraj admits he expects such "time-dependent factors" (including positive demand shocks) would affect "[a]ll website[s], including Claude.ai and lyrics website[s]." Nagaraj Report ¶119. He even cites the example of "the release of a new song from a popular artist." *Id.* While Dr. Nagaraj claims to control for such factors, his method for doing so fails to control for positive demand shocks occurring across multiple months or within a particular month.

109. Dr. Nagaraj's web traffic regressions may also produce statistically insignificant coefficients for his Claude.ai variable, even if consumers consider Claude a substitute for lyric websites. This is because Dr. Nagaraj's model does not disentangle countervailing effects: the emergence of a substitute can reduce quantity consumed for the original good, but sufficiently

large, positive demand shocks may still increase the quantity consumed despite the presence of the substitute. These competing effects could also essentially cancel out, which in this context would be associated with a statistically insignificant coefficient for the relationship between Claude.ai traffic and lyric website traffic. Thus, neither a positive nor an insignificant coefficient in Nagaraj's traffic model is sufficient to conclude that Claude is not a substitute for existing lyric websites or that Claude did not *cause* lyric website traffic to be lower than it would have in the absence of Claude.

110. Dr. Nagaraj's claim that Claude does not act as a substitute to existing lyric websites is belied by simply reviewing the sample prompt-outputs produced by Anthropic. Dr. Nagaraj cites Dr. Bray's analysis of sample prompt-outputs and Dr. Bray's opinion that lyric-related Claude usage is rare, Nagaraj Report ¶ 88, to support his claim that he "expect[s] the degree of substitution between Claude.com and online lyrics aggregators to be low," *id.* ¶ 89. Dr. Nagaraj appears to be confusing frequency with degree: the prevalence of lyric-related Claude usage does not provide any indication of whether and to what extent those lyric-related queries are a substitute for viewing those same lyrics on a lyric website.

111. Though he cites the results of Dr. Bray's findings, Dr. Nagaraj does not appear to review any of the lyric-related prompt-outputs identified by Dr. Bray. Had he done so, he would have noticed that many of the prompt-outputs flagged by Dr. Bray involve users specifically requesting song lyrics and Claude *reproducing those lyrics verbatim* in its responses. *See, e.g.*, Expert Report of Dr. Andrew Bray (Dec. 19, 2025), at ¶ 102 (reciting lyrics to, e.g., Gloria Gaynor's "I Will Survive.").

112. Despite Dr. Nagaraj's claim that "economic theory" predicts it is unlikely that Claude acts as a substitute for online lyrics aggregators, Nagaraj Report ¶ 89, the available

evidence suggests that Claude could be and was used to provide the main service offered by lyrics aggregators: supplying users with the lyrics to a song they are interested in. In the absence of Claude's reproduction of the lyrics, common sense suggests these users who specifically requested (and received) a particular Work-in-Suit's lyrics on Claude, but were otherwise unable to obtain those lyrics, would instead turn to lyric websites, which also provide the online display of lyrics. Thus, neither Dr. Nagaraj's web traffic regression analyses nor "economic theory" support his conclusions that Claude is not a substitute for lyrics websites and did not cause a negative impact on demand for lyrics websites.

113. Dr. Nagaraj's second set of regression models, purportedly evaluating Claude's effect on lyric-related Google searches relative to song-related Google searches, similarly fails to establish causality. Dr. Nagaraj claims to use an interaction term to measure the effect of Claude usage on the relationship between lyric and song searches on Google. Specifically, he states he would expect a negative, statistically significant coefficient on an interaction term between a variable indicating "lyric" Google Trends data and another variable for traffic to Claude.ai "because users would be shifting away from Google searches for lyrics (relative to their searches for songs) and relying on Claude for lyrics instead." *Id.* ¶ 142.

114. Dr. Nagaraj's models treat "song" indices as a control group, supposedly measuring the interest in listening to a particular Work-in-Suit rather than viewing its lyrics, such that the lyrics indicator variable would measure the difference in interest in lyrics against interest in audio for the same song in the same week. This is a fundamental assumption of his model that is unsupported, as his variable of interest is the interaction of the lyrics indicator variable and Claude traffic. If his song indices are not a control group, then the coefficients of both the lyrics indicator

variable and its interaction term are unreliable, as are any conclusions Dr. Nagaraj reaches based on those models.

115. Yet, Dr. Nagaraj has no basis for this central assumption in his Google Trends analyses. Dr. Nagaraj makes the following claim without citing any analysis or external source relating to the meaningfulness of the terms "song" or "lyric" in Google searches:

> First, the "song index" consists of **searches likely related to the audio of a song**, which cannot be reproduced by Claude because Claude does not output audio. The second is the "lyrics index," which consists of **searches likely related to the lyrics text of a song** (text that Plaintiffs allege could be reproduced by Claude).[53]

116. It appears that Dr. Nagaraj invented this distinction between the meaning of including the words "song" or "lyrics" after the title (or title and artist) in a user's Google search. But not only does this distinction appear to be arbitrary, it is also contradicted by even a cursory review of Google search results using the song index search terms Dr. Nagaraj created for his Google Trends analysis. For example, Figure 3 below reflects a screenshot of the first Work-in-Suit listed in Dr. Nagaraj's Table C.1 using his "[title] + [artist] + song" variation.

[53] Nagaraj Report ¶ 126 (emphasis added).

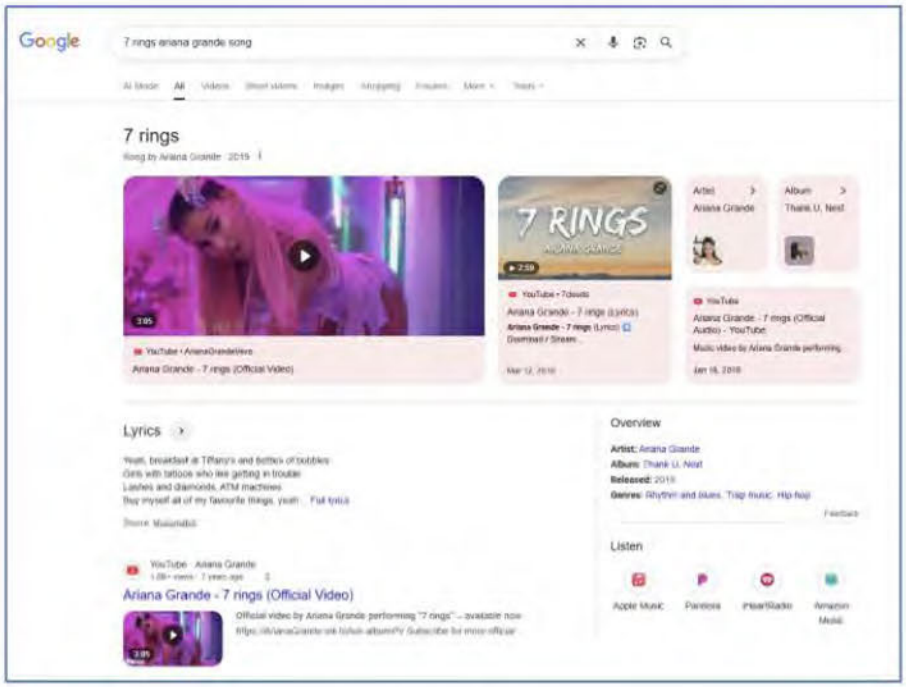


Figure 3

117. And Figure 4, below here, reflects another screenshot of the same Work-in-Suit but using Dr. Nagaraj's "[title] + song" variation. As is immediately evident, both searches yield the same Musixmatch lyrics directly embedded in the search page, meaning that both search-term variations can be readily used to search for lyrics.

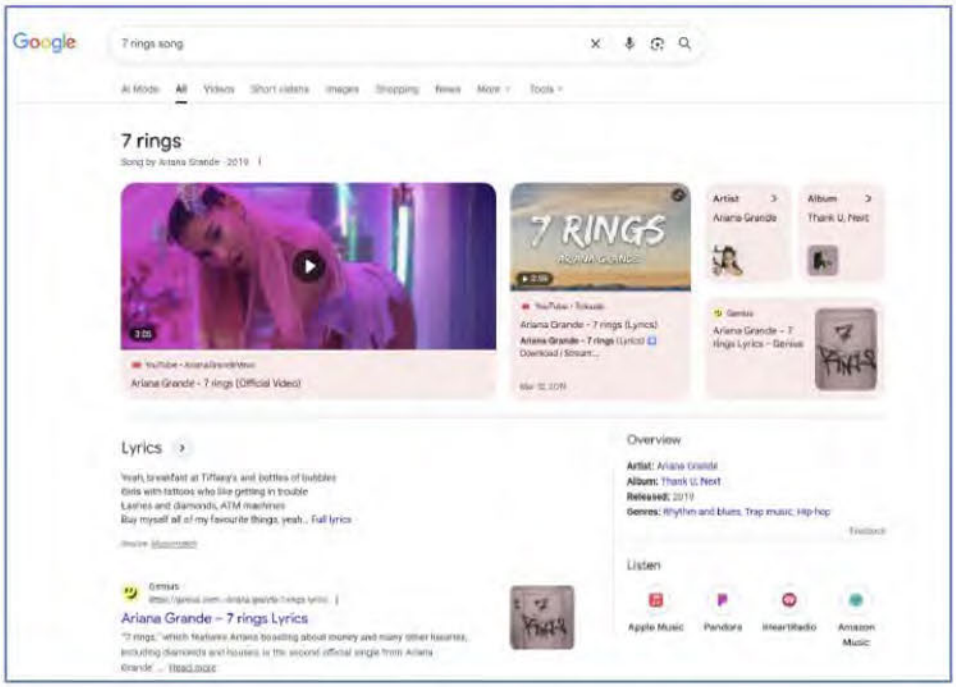


Figure 4

118. Beyond these illustrative examples, I also reviewed the top 10 search results for every search terms Dr. Nagaraj used in his construction of his Google Trends regression dataset.[54] By isolating the domain names for each of the search results, it's then possible to identify the most common websites across all the song index search terms. Seven of the lyric websites identified by Dr. Nagaraj in his web traffic analyses appear in these top "song" search results. Further, as Figure 5 shows, Genius.com is the 6th most common website across the song index search terms' top 10 results.[55] I also note that other webpages appearing in the top search results may contain some or all of the lyrics of the song being searched for.

[54] For the purposes of this analysis, the search terms are based on the "title" and "title_artist" fields produced by Dr. Nagaraj in "02_df_title_generic.csv" with the word "song" appended to each.

[55] This analysis is based on the combined set of "Title" and "Title Artist" search results with "song" appended to each.

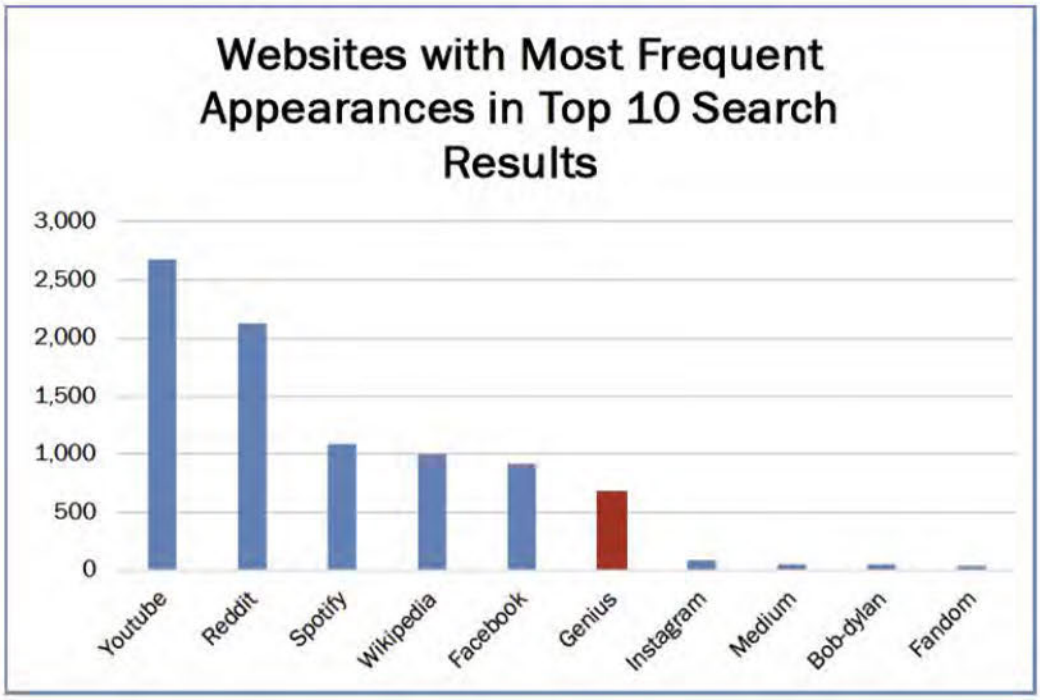


Figure 5

119. Dr. Nagaraj does not cite his professional experience or any other evidence to determine whether the absence of the specific word "lyrics" in a Google search means the user is specifically *not* looking for lyrics. As my analysis shows, lyric websites' prevalence as top search results for the song index search terms suggests that there is a high likelihood that Google Trends values in Dr. Nagaraj's song index also reflect searches containing the word "song" where the consumer then views a lyric website appearing in the top search results. As a result, Dr. Nagaraj's control group would be tainted by the very effect he is trying to measure. Without support for such a key assumption in this set of analyses, his song index is not necessarily a control group and no conclusions regarding causality can be determined from any differences between the lyric index and song index.

2. The variables of interest in Dr. Nagaraj's models are manipulated and do not reflect the conduct at issue.

120. Dr. Nagaraj's regression analyses for both web traffic and Google Trends rely on Claude.ai traffic obtained from SemRush as a key variable of interest. In his web traffic analysis,

he considers the sign and p-value of the Claude.ai variable to support his conclusion that increased Claude usage did not cause a decrease in lyric website traffic. Nagaraj Report ¶ 121. Similarly, in his Google Trends analysis, Dr. Nagaraj cites the sign and p-value of the interaction term between his lyrics indicator variable and Claude.ai traffic to conclude that increased Claude usage did not cause a decrease in Google searches for lyrics relative to searches for song audio for the Works-in-Suit. *Id.* ¶ 143.

121. As such, both his web traffic and Google Trends analyses suffer from the same set of flaws which cause his models to be unreliable. Dr. Nagaraj adopts unrealistic and unsubstantiated assumptions that (i) total Claude.ai traffic is an appropriate proxy for the true metric of interest, *lyric-related* Claude usage, and (ii) Claude's regurgitation of WIS lyrics entirely ceased after Anthropic's purported implementation of guardrails (an assumption contested by Publishers' expert Mr. Dennis, as noted above). *See Id.* ¶ 114.

122. A serious issue with Dr. Nagaraj's analyses is his use of Claude.ai traffic as a proxy for lyric-related Claude usage in his regression models despite failing to conduct any analysis to confirm whether it is an appropriate proxy. Dr. Nagaraj himself states that it his understanding that "Claude is not designed to regurgitate lyrics, and that requests related to lyrics or songs are not frequent use cases." *Id.* ¶ 62. Dr. Nagaraj also cites Dr. Bray's findings that "only 0.0059% of prompts [] sought any amount or form of regurgitation of the lyrics to the Works-In-Suit." *See id.* and "less than 0.05%" of prompts produced in the sample sought any amount or form of lyric regurgitation. *Id.* ¶ 88.

123. While Dr. Nagaraj argues that lyric-related prompts are a miniscule percentage of overall Claude usage elsewhere in his report, he still uses data on overall Claude usage in all of his regression analyses. Despite this clear logical inconsistency, Dr. Nagaraj makes no attempt to

justify the validity of using general traffic to Claude.ai or Google searches for "claude ai," rather than a specific measure of *lyric-related* Claude usage.

124. This omission is especially problematic given the potential effect on both the coefficients and statistical significance of key variables of interest in his regression model. In regression models, coefficients and statistical significance reflect the relation between an explanatory variable and the dependent variable, specifically the predicted change in the dependent variable in response to a change in that explanatory variable while holding all else equal.[56] When using a proxy for a variable of interest, it is important that the proxy move similarly to the variable of interest; otherwise the proxy's coefficient and significance will not reflect the variable of interest's true relation with the dependent variable.

125. A disconnect between the proxy and the variable of interest can invalidate statistical results based on the proxy. In any given week, overall Claude usage may increase while lyric-specific Claude prompts decrease, or vice versa, much like an individual stock's price can increase sharply while the broader stock market remains unchanged or even declines. As an example, if lyric-website traffic were to decrease in weeks when lyric-specific Claude prompts increase, yet overall Claude traffic does not move similarly, Dr. Nagaraj's models would be expected to find the coefficient of his Claude.ai traffic variable to be insignificant, or even significant and positive.

126. Because lyric-related prompts are a relatively small proportion of overall Claude prompts, even large increases in lyric-related Claude usage can be easily "overwhelmed" by uncorrelated changes in overall Claude usage. As an illustrative example, there could be 5 lyric-related prompts and 9,995 non-lyric prompts in Week 1 (a prevalence rate of 0.5%), and then lyric-

---

[56] *See* Jeffrey M. Wooldridge, INTRODUCTORY ECONOMETRICS: A MODERN APPROACH (5th ed. 2012), at p. 74.

related prompts double (an increase of 100%) in Week 2. Yet, if non-lyric prompts decreased by a mere 0.06%, the overall number of prompts would decrease from Week 1 to Week 2. As this illustrative example highlights, the disparity between total Claude prompts and lyric-related Claude prompts calls into question the reliability of any conclusions regarding lyric-related Claude usage that Dr. Nagaraj draws from interpretation of coefficients and statistical significance for variables relying on overall Claude traffic or Google searches for "claude ai."

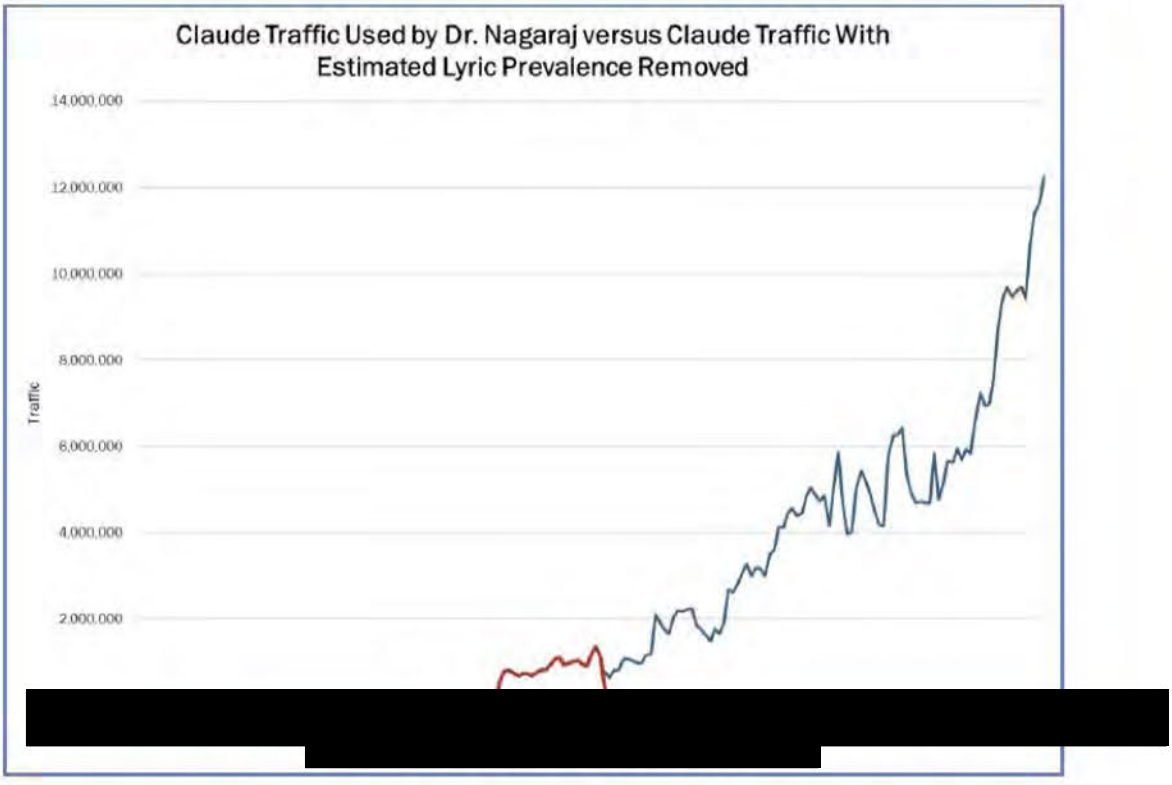
127. Dr. Nagaraj also errs in modifying Claude usage data based on the assumption that Claude's regurgitation of Works-in-Suit lyrics entirely ceased after Anthropic's purported implementation of guardrails. While Dr. Nagaraj uses Claude.ai's actual traffic for the period July 11, 2023, through [redacted], for web traffic and [redacted], for Google Trends, *see* Nagaraj Report ¶¶ 111, 135, he overwrites all values after these dates, forcing Claude.ai usage to be 0 for all dates afterward. Dr. Nagaraj claims to do so in order to account for Anthropic's purported implementation of guardrails, which he alleges "dramatically reduced the possibility that Claude.ai could substitute for song lyrics." *Id.* ¶ 111.

128. This assumption is unfounded; as Publishers' expert Joshua Dennis has shown, Claude continued regurgitating WIS lyrics after Anthropic's implementation of guardrails. *See* Expert Report of Joshua Dennis (Dec. 19, 2025) (the "Dennis Report"), at ¶ 217; *see also* Expert Rebuttal Report of Shawn Shan (Jan. 23, 2026), at ¶¶ 36, 42–43.

129. This assumption also creates a logical inconsistency in that Dr. Nagaraj's variable for measuring Claude.ai usage is actually a combination of two different metrics: overall Claude traffic prior to the guardrail implementation and alleged lyric-only Claude usage after the

guardrail.[57] Despite this modification, Dr. Nagaraj does not claim that this variable is specifically measuring lyric-related Claude traffic. Dr. Nagaraj does not even claim that his Claude usage variables are even approximating lyric-related Claude usage.

130. For the purposes of regression analysis, the post-guardrail zeroing of Claude usage variables is also notable because it appears in the data as a massive decline in Claude.ai traffic, well beyond the number of lyric-related prompts implied by the prevalence rates found in the sample prompts produced by Anthropic. Figure 6 shows the decline in Claude traffic exhibited in Dr. Nagaraj's manipulated data relative to Claude traffic modified in the post-guardrail period to remove the percentage of lyric-related prompts claimed by Dr. Bray's analysis that Dr. Nagaraj cited elsewhere in his report.



[57] Note this is based on Dr. Nagaraj's claim that Claude's guardrails fully prevented the reproduction of song lyrics.

Figure 6

131. In the primary regressions displayed in his report,[58] the Claude traffic variable is the only non-indicator independent variable (*i.e.*, the only independent variable that does not exclusively take values of 0 or 1).[59] The Claude traffic variable is also the only independent variable that varies week-to-week as the dependent variable does, and thus the only independent variable capable of explaining week-to-week variation in the dependent variable (lyrics traffic or Google Trends index values) after the dependent variable is demeaned by the fixed effects. Thus, only around 12.1% of observations in his web traffic regressions and around 9.5% in his Google Trends regressions are even relevant to the effect he is purporting to measure: weekly variations in lyric interest explained by changes in Claude usage.[60]

3. Dr. Nagaraj's use of fixed effects obscures the low explanatory power of his models and the relationship between Claude and lyric interest.

132. In both his web traffic and Google Trends regression analyses, Dr. Nagaraj includes fixed effects intended to control for group-level differences. In Dr. Nagaraj's web traffic analyses, he uses fixed effects (or "FE") to control for differences across lyrics websites ("site fixed effects") and separate fixed effects for each unique combination of month and year ("month-year fixed effects") in his data to control for what he dubs "time-dependent" factors. *See* Nagaraj Report ¶ 119. In his Google Trends analyses, he again adopts month-year fixed effects, as well as fixed effects to control for differences across Works-in-Suit ("song fixed effects").

---

[58] Nagaraj Report, Tables 1 and 2.

[59] While the interaction term varies by week as well, this is because it is the product of the Claude traffic variable and an indicator variable that only takes values 0 or 1.

[60] *See* Nagaraj Report, Tables 1 and 2, and back-up data.

133. I first address the issues associated with Dr. Nagaraj's use of fixed effects in his web traffic regressions, with which he purports to control for "time-dependent factors" and for differences in web traffic across the 26 lyrics websites he considers. I find that Dr. Nagaraj purported controls for "time-dependent factors" potentially obscure the true relationship between Claude usage and lyric website traffic. I also find that he inappropriately applies website fixed effects in a manner which obscures the impact of Claude usage on the overall market for the online display of lyrics.

134. Rather than considering what control variables are relevant in a model intended to estimate demand for lyric websites or lyric-related Google searches, Dr. Nagaraj seems to view his month-year fixed effects as a catchall method of controlling for any and all factors that might vary across time outside of Claude.ai usage. Dr. Nagaraj cites examples of time-dependent factors such as "the release of a new song from a popular artist, or the season, all of which may impact web traffic" and which he claims are controlled for through his "month-year fixed effect." *Id.* However, Dr. Nagaraj does not provide any supporting analysis to determine that time-dependent factors were significant nor that month-year fixed effects were the appropriate measure to address such factors, as compared to other approaches less susceptible to overfitting, such as seasonal indicator variables.[61]

135. This method of controlling for time-independent factors is inappropriate in this setting for several reasons. First, it can fail to capture the true relation between Claude usage and lyric website traffic, potentially attributing monthly changes in lyric web traffic to the numerous

[61] Seasonal dummy variables are used to control for specific seasonal patterns (regular changes in the summer, or during the Christmas holiday, for example), whereas month-year dummy variables assume each month in the dataset is inherently unique—and may absorb variation associated with the effect you are trying to measure.

month-year variables even if the change in lyric web traffic was actually driven by a change in Claude usage. Second, it fails to account for "time-dependent factors" which span across months or persist for less than a month. Third, it obscures meaningful information about movements in both Claude and lyric traffic, with the obvious example being one Dr. Nagaraj himself cites: the release of a new song from a popular artist. If Dr. Nagaraj believes Claude and lyrics website traffic would both increase in response to the release of a new song, it suggests demand for Claude and lyrics website traffic is related in some way. Yet, according to Dr. Nagaraj, his models would instead attribute this pattern to his month-year fixed effects.

136. Dr. Nagaraj's excessive use of fixed effects also strips out any long-term trends evident in the data. Month-year fixed effects are functionally the same as subtracting the average value in a particular month-year from every dependent and independent variable value in that month-year. Even if web traffic doubled from one month to the next, both months' average web traffic would be 0 after being demeaned. Mostly Harmless Econometrics, which Dr. Nagaraj cites, *see* Nagaraj Report n.149, n.176, echoes this concern, observing: "A variant on the measurement-error problem arises from that fact that the differencing and deviations-from-means estimators used to control for fixed effects typically remove both good and bad variation. In other words, these transformations may kill some of the omitted-variables-bias bathwater, but they also remove much of the useful information in the baby - the variable of interest."[62]

137. A further concern is the reliance on a panel data approach in his regression analysis. As Figure 7 shows, the vast majority of web traffic in Dr. Nagaraj's data goes to one site, with 74.3% of the web traffic attributable to Genius. The next largest site, AZLyrics, makes up only

---

[62] *See* Joshua D. Angrist and Jörn-Steffan Pischke, MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST'S COMPANION (2008), at p. 168.

10% of the web traffic in Dr. Nagaraj's dataset, while 19 of the 26 websites each constitute less than 1% of the web traffic in his dataset. Given that Dr. Nagaraj purports to measure the effect of Claude usage on the market for lyric websites overall, I see no economic or statistical basis for why he adopts a panel data approach measuring each website's traffic across time, rather than aggregating the lyric websites' traffic. Further, I note that several of the small lyrics websites he includes appear to have highly volatile weekly traffic, making it unlikely that the only non-FE independent variable, Claude traffic, would be able to explain this volatility.

SemRush Traffic by Lyric Website

Lyric Website	Total Traffic	Pct. of Total
genius.com	2,976,259,607	74.3%
azlyrics.com	400,187,078	10.0%
lyricstranslate.com	122,490,895	3.1%
lyrics.com	113,951,555	2.8%
musixmatch.com	105,110,875	2.6%
songmeanings.com	90,085,330	2.2%
songlyrics.com	53,727,588	1.3%
lyricsondemand.com	30,214,251	0.8%
multitracks.com	20,267,933	0.5%
lyricfind.com	13,828,428	0.3%
lyricsmode.com	11,086,097	0.3%
lyricsfreak.com	10,857,609	0.3%
eminem.com	10,078,183	0.3%
bobdylan.com	8,217,961	0.2%
darklyrics.com	7,460,631	0.2%
allthelyrics.com	6,680,675	0.2%
elyrics.net	5,342,650	0.1%
letssingit.com	4,214,200	0.1%
lyricsbox.com	3,571,518	0.1%
lyricsmania.com	3,223,260	0.1%
worshiponline.com	3,045,115	0.1%
lyricstraining.com	1,885,210	0.0%
lyricsify.com	1,155,979	0.0%
woodyguthrie.org	788,937	0.0%
lyricsdepot.com	662,191	0.0%
seekalyric.com	35,921	0.0%
Total	4,004,429,677	100.0%

Figure 7

138. Much like his web traffic analysis, Dr. Nagaraj's regression analyses of Google Trends data suffer from a flawed implementation of fixed effects for this setting. First, while Dr. Nagaraj's use of song fixed effects appears to mirror his use of website fixed effects for the

dependent variable in his web traffic models, I find that the nature of Google Trends data makes this control unnecessary and redundant. Dr. Nagaraj states that "I include song-fixed effects in my model to account for any characteristics that are unique to individual songs" (Nagaraj Report ¶ 138), whereas in his web traffic analysis he claims to control for "the vast variation in popularity across different websites" (*id.* ¶ 118), presumably equating popularity and traffic levels. However, Google Trends data already normalizes search volumes for each query on a 0–100 scale,[63] which would account for differences in popularity across different songs. Furthermore, his models purport to measure each WIS's lyric index against a benchmark of the same WIS's song index, which should also account for differences between songs. Despite these obvious characteristics of the data he relies on and the model specification he adopts, Dr. Nagaraj cites no supporting analysis and offers no opinion as to how he determined song fixed effects were necessary to include in these models.

139. Similar to his song fixed effects, Dr. Nagaraj's use of a month-year fixed effect in his Google Trends regressions is also redundant and unnecessary due to his use of a benchmark song index. Economic theory would suggest that both the song index and lyric index would be expected to respond similarly to "potential trends impacting search interest for songs over time" for which Dr. Nagaraj purports to control with his month-year fixed effects in his Google Trends analyses. *See* Nagaraj Report ¶ 138. In fact, Dr. Nagaraj also claims that "the song index is an appropriate control to include in my model in order to account for general interest in the song." *Id.* Dr. Nagaraj cites no supporting analysis and offers no opinion as to how he determined that the song index benchmark, song fixed effects, and month-year fixed effects were *all* necessary to include in his models.

---

[63] GOOGLE, *Google Trends*, https://trends.google.com/trends/ (last visited Jan. 22, 2026).

140. In sum, based on my review of Dr. Nagaraj's data, models, and purported rationale for his methodology, I find that Dr. Nagaraj fails to adhere to reliable economic principles and methods in his empirical analysis, rendering his conclusions from these models unreliable and prone to misleading conclusions.

141. I also note that these models only analyze the very short-run impact of Anthropic's alleged regurgitation of lyrics on visits to lyric sites. They do nothing to capture potential longer-term harm on Publishers' licensing markets from the use of unlicensed human lyrics to create a tool that holds the potential to replace human artists. *See* Smith Report ¶ 66.

III. REBUTTAL TO MS. DAVIS'S OPINIONS

142. Ms. Davis states that she and her consulting practice, Stout Risius Ross, LLC, were retained by Anthropic's outside counsel to "address and analyze certain elements of a copyright damages claim for which the Music Publishers would be entitled to a remedy of Anthropic's profits." *See* Julie Davis Report (Dec. 19, 2025) (herein "Davis Report"), at p. 3. She says that she was also "asked to analyze the financial data produced by the Plaintiffs, as well as other information, as it relates to Factor 4 of the fair use defense, which is 'the effect of the use upon the potential market for or value of the copyrighted work.'" Davis Report at p. 3.

143. Ms. Davis's report "outline[s] the apportionment factors that [she] expect[s] to use to calculate the portion of Anthropic's profits that could potentially be attributable to Anthropic's alleged infringement[.]" *Id.* at p. 4. I have identified multiple problems with those "apportionment factors" and her "observations related to Factor 4 of the fair use defense." *Id.*

144. First, Ms. Davis and Dr. Nagaraj disagree about the relevant market for purposes of their respective "factor four" analyses. As discussed above, Dr. Nagaraj claims that the "correct market to analyze" for purposes of his factor four analysis is the "market for all of the data needed to train a general-purpose LLM," Nagaraj Report ¶ 8, while Ms. Davis uses a much broader

"market for licensing the lyrics of the Asserted Works" in connection with her own factor four analysis. Davis Report at p. 51. While I disagree with her ultimate conclusion that there has been no harm to Publishers, Ms. Davis has the better market model for purposes of conducting a "factor four" analysis in this case because her model directly addresses the "potential market for or value of the ***copyrighted work***" without any of the distortions Dr. Nagaraj added. 17 U.S.C. § 107(4). In any case, the fact that Ms. Davis uses a wholly different (and far simpler) market for her factor-four analysis underscores my belief that Dr. Nagaraj's principal and alternative markets are both incorrect.

145. Second, with respect to her apportionment analysis, Ms. Davis erroneously relies on "total" tokens used to train Claude as her basis for apportioning profit. *See* Davis Report at pp. 42–45. As noted above, *see* ¶ 70, not all training data is valued equally, as some datasets are more impactful on AI performance than others. Consequently, and as I discuss above, it does not make sense to apportion Anthropic's profits based on "the portion of tokens associated with [Publishers' lyrics] as a percentage of the total tokens used to train Claude" because such an approach will not take into account specific value derived from those lyrics, including with respect to specific Claude use cases, like songwriting or lyric generation, which may serve as a draw for particular classes of Anthropic customers. *See* Smith Report ¶¶ 201–02. Further, as discussed above, the value of any particular dataset for AI training is relative to what the training data already includes, as adding uniqueness and diversity is especially valuable. *See* above ¶ 70; *see also* P. Sangani 30(b)(1) Tr. 99:22–100:18 (explaining that "how much data already exists as part of constituent data mixes" is a factor in how Anthropic values training data). Any apportionment analysis must account for the specific value derived from the copyrighted works at issue and cannot simply assume that all training tokens can be valued equally.

146. Third, Ms. Davis's "Claude Output" analysis is flawed because it does not account for any value Anthropic derived from using Publishers' copyrighted lyrics in training the Claude models to create outputs *other than* regurgitated lyrics. *See* Davis Report at pp. 45–47. For example, Anthropic derives value from the use of Claude to generate "new" song lyrics that compete with Publishers' lyrics. *See* Smith Report ¶¶ 201–02.

147. Ms. Davis's output discussion also fails to account for and does not acknowledge the fact that the 5-million prompt/output sample analyzed by Anthropic's expert Dr. Bray is statistically representative of a much larger sample of 455 million prompt/output pairs across a 6-month period. *Compare* Davis Report at pp. 45–47 *with* Smith Report ¶ 86 *and* ECF No. 377. Thus, the numbers she relies on from Dr. Bray's analysis must be extrapolated to account for the full universe of Anthropic prompts/outputs over the relevant time period. And last, her claim that "there were not outputs providing lyrics to the Asserted Works after [redacted]" (when Anthropic implemented guardrails) is contested by Publishers' expert Mr. Dennis. *Compare* Davis Report at p. 46 n.280 (and accompanying text) *with* Dennis Report at ¶ 217.

148. Fourth, in connection with her "factor four" analysis, Ms. Davis attempts to determine whether "there has been harm to the market for licensing the lyrics of the Asserted Works" solely by reference to recent revenue data produced by Plaintiffs. *See* Davis Report at pp. 48–53. This is a narrow method of analysis that does not account for any "but-for" worlds or external factors that may cause noise in the data. Similarly, this analysis does not account for the fact that not all lyric-display contracts pay royalties based on the number of views, which means Publishers' revenue data will not reflect decreased revenue derived from those agreements until the next round of contract renewals. For example, [redacted]. *See* Smith Report ¶ 163. As a result, Ms. Davis's

analysis does not account for any loss in Publishers' negotiating power when renewing or extending existing agreements. Similarly, Ms. Davis's analysis fails to account for any delayed or long-term impact of the kind I discuss in my initial report. *See* Smith Report ¶¶ 65–67.

149. Ms. Davis ends her factor four observations by claiming, without any evidence, that generative AI may benefit the market for licensing lyrics. Ms. Davis does not attempt to quantify or account for any supposed benefits beyond noting generic statements Publishers have made in deposition or publicly to the United States Copyright office. Moreover, Ms. Davis also does nothing to link any of the supposed benefits she claims will flow from generative AI to the use of lyrics *as training data* or the market for lyrics as training data: the supposition that generative AI might have beneficial uses in the licensing market for lyrics is irrelevant to the "effect of ***the use*** upon the potential market for or value of the copyrighted work," which is supposed to be the subject of her analysis. Davis Report at p. 3.

150. Accordingly, it is my opinion that, while Ms. Davis uses the better market for her factor-four analysis as compared to Dr. Nagaraj, her observations are either flawed, unsupported, or irrelevant to "the effect of [Anthropic's] use upon the potential market for or value of [Publishers'] copyrighted work[s]." *Id.* (citing 17 U.S.C. § 107(4). Additionally, it is my opinion that her apportionment analysis is flawed because of its methodological assumptions.

Rebuttal Exhibit A

151. There is no perfect way to use public sources to measure the size of an AI licensing market for copyrighted content. Nonetheless, one useful starting point is the Copyright Alliance list of "AI Licensing by AI Companies" that, as of January 19, 2026, lists 211 licensing deals between AI companies and rightsholders.[64] It is important to note that the 211 number likely represents a fraction of the total number of deals in the marketplace—and also a fraction of the deals represented in this list. For example, the Copyright Alliance's page only lists 14 individual rightsholders from the "approximately 4,000 publishers" reportedly represented by Factiva/Dow Jones, and includes only 55 rightsholders among the "over 700 news publications and other rightsholders" reportedly licensed by Prorata.ai (https://prorata.ai/). That said, my point is not to provide a complete catalog of deals, but rather to point to evidence of a large, active, and growing licensing market between rightsholders and Generative AI companies.

152. As there is no perfect way to use public sources to measure the size of this market, there is no perfect way to measure the growth of this market over time. Nonetheless, the Copyright Alliance list again provides a useful starting point to illustrate the growth of this market. To do this, I followed the URL link for each of the deals listed in the Copyright Alliance list and used the date on that link as the (admittedly latest) date of the license.

153. In cases where there were no links, I attempted to find a reasonable approximate license date from public sources. For the 14 Factiva/Dow Jones licensing deals listed on the page

---

64 *See* COPYRIGHT ALLIANCE, *AI Licensing by AI Companies*, https://copyrightalliance.org/artificial-intelligence-copyright/licensing/ai-companies/ (last visited Jan. 22, 2026).

I used November 13, 2024, the date in this webpage announcing the launch of the service.[65] For the deal between Exponential View and ProRata I used November 20, 2024, the date in this press release announcing the deal.[66] Similarly, for the deal between Mediahuis and ProRata I used November 20, 2024, the date in this press release announcing the deal.[67] For all of the licenses with Beatoven.ai, I used August 27, 2025.[68] And in the case of Wirestock.io and vAIsual I used August 2, 2023, the date in this announcement.[69]

154. In several cases the Copyright Alliance linked to an amicus brief from the Association of American Publishers in the *Kadrey v Meta* case which listed "AI licensing deals for textual works of which AAP is aware, either directly from its members or through public

---

[65] *See* DOW JONES, *Dow Jones Launches Factiva Smart Summary,* https://www.dowjones.com/press-room/dow-jones-launches-factiva-smart-summary/ (last visited Jan. 23, 2026).

[66] *See* FINANCIAL CONTENT, *Major UK Media Organizations Join ProRata's Movement to Credit and Compensate Content Owners in the Age of Generative AI,* https://markets.financialcontent.com/stocks/article/bizwire-2024-11-20-major-uk-media-organizations-join-proratas-movement-to-credit-and-compensate-content-owners-in-the-age-of-generative-ai (last visited Jan. 23, 2026).

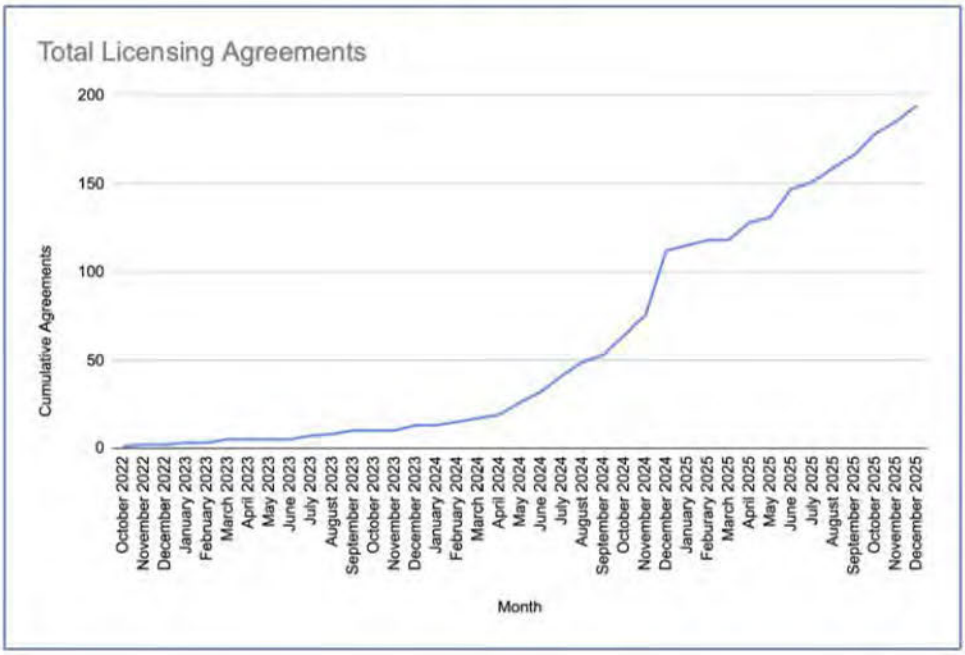
[67] *See* BUSINESS WIRE, *Major UK Media Organizations Join ProRata's Movement to Credit and Compensate Content Owners in the Age of Generative AI,* https://www.businesswire.com/news/home/20241119176784/en/Major-UK-Media-Organizations-Join-ProRatas-Movement-to-Credit-and-Compensate-Content-Owners-in-the-Age-of-Generative-AI (last visited Jan. 23, 2026).

[68] *See* MUSIC BUSINESS WORLDWIDE, *Beatoven.ai Launches 'Fully Licensed' Music AI Model – With Artist Payouts,* https://www.musicbusinessworldwide.com/beatoven-ai-launches-fully-licensed-music-ai-model-with-artist-payouts/ (last visited Jan. 23, 2026).

[69] *See* VAISUAL, *vAIsual and Wirestock.io Partner to Offer 9 Million Commercial Stock Images and Video for AI Training,* https://vaisual.com/vaisual-and-wirestock-io-partner-to-offer-9-million-commercial-stock-images-and-video-for-ai-training/ (last visited Jan. 23, 2026).

reports."[70] In that case, I used the date of the filing (April 11, 2025). I also found one broken link for the deal between Alex Springer and OpenAI[71] and used December 12, 2023, the date that is shown in the Google search result for that URL, which is within a few days of the date in this separate announcement.[72]

155. From this, I created a chart of the number of cumulative licenses over time listed among the Copyright Alliance list in the Figure below.

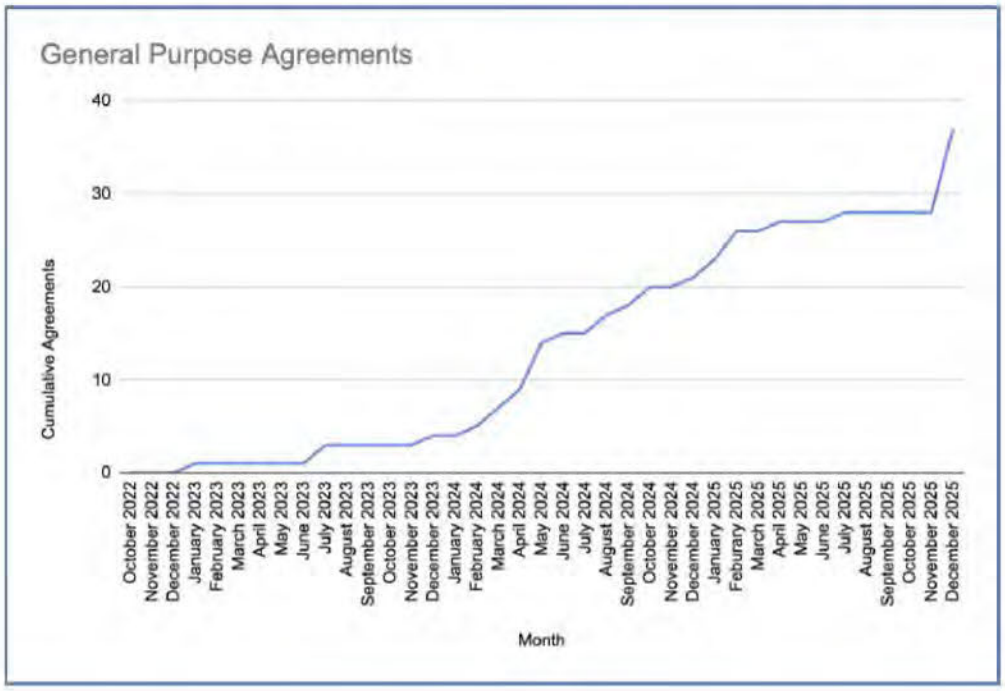


[70] Brief of Amicus Curiae Association of American Publishers in Support of Plaintiffs' Motion for Partial Summary Judgment and Opposition to Defendant's Motion for Partial Summary Judgment, at 12, *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC (N.D. Cal. Apr. 11, 2025).

[71] *See* AXEL SPRINGER, *Nothing to See Here*, https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism (last visited Jan. 23, 2026).

[72] *See* AITHORITY, *Axel Springer and OpenAI Partner to Deepen Beneficial Use of AI in Journalism*, https://aithority.com/machine-learning/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism (last visited Jan. 23, 2026).

156. I then limited my analysis to just licenses reported to be signed by one of the four General Purpose LLMs: Anthropic, Google, Meta, and OpenAI. The cumulative number of licenses in the Copyright Alliance list for these General Purpose LLMs is shown in the table below.



REBUTTAL EXPERT REPORT OF MICHAEL SMITH

APPENDIX 1: DOCUMENTS CONSIDERED

LITIGATIONS

Case Name	Case No.	Court
Alter v. OpenAI, Inc.	No. 1:23-cv-10211-SHS	N.D. Cal.
Authors Guild v. OpenAI, Inc.	No. 1:23-cv-08292	N.D. Cal.
Chabon v. OpenAI, Inc.	No. 3:23-cv-04625	N.D. Cal.
DENIAL v. OpenAI, Inc.	No. 3:25-cv-05495	N.D. Cal.
Huckabee v. Meta Platforms, Inc.	No. 24-cv-773-GBW	D. Del.
In re Google Generative AI Copyright Litigation	No. 5:23-cv-03440-EKL	N.D. Cal
In re: OpenAI, Inc. Copyright Infringement Litigation	No. 1:25-md-3143	S.D. Cal.
Intercept Media, Inc. v. OpenAI, Inc.	No. 1:24-cv-01515-JSR	S.D.N.Y.
Kadrey et al. v. Meta Platforms, Inc	No. 3:23-cv-03417-VC	N.D. Cal.
Millette v. OpenAI, Inc.	No. 5:24-cv-04710	N.D. Cal.
Raw Story Media, Inc. v. OpenAI, Inc.	No. 1:24-cv-01514	N.D. Cal.
Silverman v. OpenAI, Inc.	No. 3:23-cv-03416	N.D. Cal.
The Center for Investigative Reporting, Inc. v. OpenAI, Inc.	No. 1:24-cv-04872	S.D.N.Y.
The New York Times v. OpenAI, Inc., et al.	No. 1:23-cv-11195	S.D.N.Y
Tremblay v. OpenAI, Inc.	No. 3:23-cv-03223-AMO	N.D. Cal.

DISCOVERY, PLEADINGS, AND PRODUCED DOCUMENTS

Action, Discovery, or Pleading
Expert Report of Anthropic Expert Jeff Bilmes (December 19, 2025)
Expert Report of Anthropic Expert Dr. Andrew Bray (Dec. 19, 2025)
Expert Report of Anthropic Expert Julie L. Davis (December 19, 2025)
Expert Report of Publishers' Expert Joshua Dennis (Dec. 19, 2025)
Expert Report of Anthropic Expert Abhishek Nagaraj (December 19, 2025)
Expert Rebuttal Report of Publishers' Expert Shawn Shan (Jan. 23, 2026)
Expert Report of Publishers' Expert Shawn Shan (Dec. 19, 2025)
Expert Report of Publishers' Expert Michael Smith (December 19, 2025)
Expert Rebuttal Report of Publishers' Expert Benjamin Zhao (Jan. 23, 2026)

Deposition Transcripts and Exhibits
Dario Amodei 30(b)(1) Deposition Transcript and Exhibits (January 8, 2026)
Dawn Drain 30(b)(1) Deposition Transcript and Exhibits (November 12, 2025)
Kent Draughon 30(b)(6) Deposition Transcript and Exhibits (November 5, 2025)
Pranay Sangani 30(b)(1) Deposition Transcript and Exhibits (November 10, 2025)
Tom Turvey 30(b)(6) Deposition Transcript and Exhibits (November 5, 2025)

Bates-Stamped Documents
Anthropic_0000228080
Anthropic_0000229152
Anthropic_0000231226
Anthropic_0000429496
Anthropic_0000441006
Anthropic_0000462964
Anthropic_0000597354

INTERNET AND ACADEMIC SOURCES

1. AITHORITY, *Axel Springer and OpenAI Partner to Deepen Beneficial Use of AI in Journalism,* https://aithority.com/machine-learning/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism/ (last visited Jan. 23, 2026).
2. ASCAP, *About ASCAP*, https://www.ascap.com/about-us (last visited Jan. 23, 2026); BMI, *About BMI*, https://www.bmi.com/about (last visited Jan. 23, 2026).
3. ASCAP, *Comments Before the United States Copyright Office in the Matter of Artificial Intelligence and Copyright* (October 30, 2023) (available at https://www.ascap.com/~/media/files/pdf/articles/2023/ascap-comments-to-us-copyright-office-on-generative-ai.pdf/) (last visited Jan. 22, 2026).
4. Axel Springer, *Nothing to See Here*, https://www.axelspringer.com/en/ax-press-release/axel-springer-and-openai-partner-to-deepen-beneficial-use-of-ai-in-journalism (last visited Jan. 23, 2026).
5. Brief of Amicus Curiae Association of American Publishers in Support of Plaintiffs' Motion for Partial Summary Judgment and Opposition to Defendant's Motion for Partial Summary Judgment, at 12, *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417-VC (N.D. Cal. Apr. 11, 2025).
6. BUSINESS WIRE, *Major UK Media Organizations Join ProRata's Movement to Credit and Compensate Content Owners in the Age of Generative AI,* https://www.businesswire.com/news/home/20241119176784/en/Major-UK-Media-Organizations-Join-ProRatas-Movement-to-Credit-and-Compensate-Content-Owners-in-the-Age-of-Generative-AI (last visited Jan. 23, 2026).

7. COPYRIGHT ALLIANCE, *AI Licensing by AI Companies*, https://copyrightalliance.org/artificial-intelligence-copyright/licensing/ai-companies/ (last visited Jan. 22, 2026).
8. DOW JONES, *Dow Jones Launches Factiva Smart Summary,* https://www.dowjones.com/press-room/dow-jones-launches-factiva-smart-summary/ (last visited Jan. 23, 2026).
9. FINANCIAL CONTENT, *Major UK Media Organizations Join ProRata's Movement to Credit and Compensate Content Owners in the Age of Generative AI,* https://markets.financialcontent.com/stocks/article/bizwire-2024-11-20-major-uk-media-organizations-join-proratas-movement-to-credit-and-compensate-content-owners-in-the-age-of-generative-ai (last visited Jan. 23, 2026).
10. GOOGLE, *Gemini*, https://gemini.google.com (last visited Jan. 22, 2026).
11. GOOGLE, *Google Trends*, https://trends.google.com/trends/ (last visited Jan. 22, 2026).
12. Joshua D. Angrist and Jörn-Steffan Pischke, MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST'S COMPANION (2008), at p. 168.
13. Jeffrey M. Wooldridge, INTRODUCTORY ECONOMETRICS: A MODERN APPROACH (5th ed. 2012), at p. 74.
14. Kate Knibbs, *Here's Proof You Can Train an AI Model Without Slurping Copyrighted Content*, WIRED (Mar. 20, 2024) (available at https://www.wired.com/story/proof-you-can-train-ai-without-slurping-copyrighted-content/) (last visited Jan. 23, 2026).
15. META, *Llama 4: Leading Intelligence*, https://www.llama.com/ (last visited Jan. 22, 2026).
16. THE MLC, *About The MLC*, https://www.themlc.com/our-story (last visited Jan. 23, 2026); SOUNDEXCHANGE, *About SoundExchange*, https://www.soundexchange.com/who-we-are/#about-us (last visited Jan. 23, 2026).
17. MUSIC BUSINESS WORLDWIDE, *Beatoven.ai Launches 'Fully Licensed' Music AI Model – With Artist Payouts,* https://www.musicbusinessworldwide.com/beatoven-ai-launches-fully-licensed-music-ai-model-with-artist-payouts (last visited Jan. 23, 2026).
18. Robert S. Pindyck & Daniel L. Rubinfeld, MICROECONOMICS, 8th ed. (2013).
19. Stella Jia & Abhishek Nagaraj, *Cloze Encounters: The Impact of Pirated Data Access on LLM Performance*, NAT'L BUREAU OF ECON. R., Working Paper No. 33598 (Mar. 2025) (available at https://www.nber.org/system/files/working_papers/w33598/w33598.pdf) (last visited Jan. 22, 2026); *see also* Anthropic_0000597354, at 59 (copy of same produced by Anthropic).
20. OPENAI, *ChatGPT*, https://chatgpt.com/overview/ (last visited Jan. 22, 2026).
21. Pierre Azoulay, Joshua L. Krieger, & Abhishek Nagaraj, *Old Moats for New Models: Openness, Control, and Competition in Generative AI*, NAT'L BUREAU OF ECON. R., Working Paper No. 32474, at p. 16 (Rev. July 2024) (available at https://www.nber.org/system/files/working_papers/w32474/w32474.pdf)(last visited Jan. 21, 2026).

22. Ravikash Bakolia, *Anthropic CEO Talks Potential IPO, Gemini and AI tech at Davos*, SEEKING ALPHA (Jan. 20, 2026) (available at https://seekingalpha.com/news/4540510-anthropic-ceo-talks-potential-ipo-gemini-and-ai-tech-at-davos) (last visited Jan. 22, 2026).
23. RESEARCH INFORMATION, *CCC Launches AI Systems Training License* (Mar. 5, 2025) (available at https://www.researchinformation.info/news/ccc-launches-ai-systems-training-license/) (last visited Jan. 22, 2026).
24. VAISUAL, *vAIsual and Wirestock.io Partner to Offer 9 Million Commercial Stock Images and Video for AI Training,* https://vaisual.com/vaisual-and-wirestock-io-partner-to-offer-9-million-commercial-stock-images-and-video-for-ai-training/ (last visited Jan. 23, 2026).
25. *Watch: Anthropic CEO Dario Amodei from World Economic Forum | WSJ*, YOUTUBE (Jan. 21, 2026) (available at https://www.youtube.com/watch?v=K7F6ohcBJus) (last visited Jan. 22, 2026); *see also* Keach Hagey, *Anthropic CEO Says Government Should Help Ensure AI's Economic Upside Is Shared*, WALL ST. J. (Jan. 20, 2026) (available at https://www.wsj.com/tech/ai/anthropic-ceo-says-government-should-help-ensure-ais-economic-upside-is-shared-1eab376c?st=38ZMHf).
26. WIKIPEDIA, *ChatGPT* (https://en.wikipedia.org/wiki/ChatGPT) (last visited Jan. 23, 2026).
27. WIKIPEDIA, *Performance rights organisation*, https://en.wikipedia.org/wiki/Performance_rights_organisation (last visited Jan. 23, 2026).