EXHIBIT 3

(FILED UNDER SEAL)

HIGHLY CONFIDENTIAL
ATTORNEYS' EYES ONLY
SOURCE CODE

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

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

EXPERT REPORT OF JEFFREY BILMES, PH.D.

Signed in Amsterdam, Netherlands, on December 19, 2025



Jeffrey Bilmes, Ph.D.

LLaMA-7B model was trained on roughly one trillion tokens;[604] LLaMA-2 models were trained on approximately two trillion tokens.[605] Meta's LLaMA 3 models, released in 2024, were pre-trained on a dataset consisting of over 15 trillion multilingual tokens.[606] Each iteration of Claude has similarly increased the token scale: Claude 2 trained on [redacted] tokens, Claude 3 trained on [redacted] tokens, and Claude 3.5 Sonnet trained on [redacted] tokens.[607] Anthropic anticipates that improvements to future models will require [redacted] of additional tokens.[608]

328. Against this scale, the Plaintiffs' 499 song lyrics (if assumed to have appeared once in the training data) represent an immeasurably small relative fraction of Claude's total training corpus. Since each song has approximately **429 tokens** on average, the combined contribution of all 499 works constitutes **213,891 tokens**[609]—[redacted] used for Claude 3.[610] At this magnitude, the effect of Plaintiffs' lyrics on Claude's learned parameters or outputs is extremely likely to be near zero, if present at all.

---

[604] Touvron, Hugo, Thibaut Lavril, Gautier Izacard, et al. "LLaMA: Open and Efficient Foundation Language Models." arXiv:2302.13971. Preprint, arXiv, February 27, 2023. https://doi.org/10.48550/arXiv.2302.13971 (listing 1.0T tokens for the 6.7B parameter LLaMA model).

[605] Touvron, Hugo, Louis Martin, Kevin Stone, et al. "Llama 2: Open Foundation and Fine-Tuned Chat Models." arXiv:2307.09288. Preprint, arXiv, July 19, 2023. https://doi.org/10.48550/arXiv.2307.09288 ("We trained on 2 trillion tokens of data as this provides a good performance–cost trade-off, up-sampling the most factual sources in an effort to increase knowledge and dampen hallucinations.").

[606] "Introducing Meta Llama 3: The Most Capable Openly Available LLM to Date," Meta AI, April 18, 2024, https://ai.meta.com/blog/meta-llama-3/ ("Llama 3 is pretrained on over 15T tokens that were all collected from publicly available sources…over 5% of the Llama 3 pretraining dataset consists of high-quality non-English data that covers over 30 languages.").

[607] Declaration of Jared Kaplan in Support of Defendant's Opposition to Plaintiffs' Renewed Motion for Preliminary Injunction ("Kaplan Declaration") (Aug. 22, 2024), Dkt. 212-5. 8:20-8:22 ("All of the Claude models were trained by first converting the text of the materials in the training corpus into tokens. Each version was trained using trillions of tokens: [redacted].").

[608] Declaration of Jared Kaplan in Support of Defendant's Opposition to Plaintiffs' Renewed Motion for Preliminary Injunction ("Kaplan Declaration") (Aug. 22, 2024), Dkt. 212-5. 8:20-8:22 ("Future models will require even more data to improve upon performance. We expect that subsequent models will also require [redacted] of tokens to continue to show performance gains.").

[609] See **Appendix C** for the computation of token counts for the Plaintiffs' lyrics.

[610] 213,891 tokens is, for example, 0.00002139% of one trillion tokens. It is [redacted].

329. As reference for the scale between [redacted] and 213,891 tokens,[611] if 500 tokens were printed on a standard 8x11 sheet of paper, and the resulting papers were stacked on top of each other, the [redacted]. By comparison, Plaintiffs' 213,891 tokens would fit in a single ~1.71-inch-tall ream of printer paper.[612]

Figure 39: Token Scale Magnitude



330. To the extent that it is helpful for Claude to understand music, lyrics, and the structure of songs as part of Anthropic's goal of creating a versatile, general-purpose model, there are ample data sources in Claude's training corpus—other than the song lyrics themselves—where this information is available. These alternative sources, such as books, the internet, and open-domain lyrics, would suffice to enable Claude to understand those concepts.[613]

---

[611] See **Appendix C** for the computation of token counts for the Plaintiffs' lyrics.

[612] Assuming a fixed 500 tokens per page on 8.5x11 paper, which is ~0.004 inches thick. [redacted]. 213,891 tokens / 500 tokens per page = ~428 pages. 428 x 0.004 inches per page = ~1.71 inches.

[613] Deposition of Ben Mann 30(b)(6) (Corporate Designee) Transcript, with Exhibits (Nov. 20, 2025), 114:10-115:10 ("So taking a step back, I believe that even if we didn't have the lyrics themselves, there are enough -- there's enough content in all the different kinds of data that we collect that is So taking a step back, I believe that even if we didn't have the lyrics themselves, there are enough -- there's enough content in all the different kinds of data that we collect that is about songs, about lyrics, about the process of creating songs, that Claude would continue to have a strong understanding of what song lyrics are, and that there are enough analogs to song lyrics, such as poems, that are

331. Research on model scaling shows that as datasets grow, the impact of any single data point on a model's performance decreases. Studies in machine learning demonstrate that a data point's influence tends to decrease predictably with dataset size in a log-linear manner, which means that every time the dataset size increases by a certain factor (like doubling), the influence of each single data point drops by about the same amount (i.e., some step size).[614] In other words, for a model trained on billions or trillions of tokens, any individual sequence contributes negligibly, if at all, to overall behavior. Research into neural networks has found that neural networks learn concepts in a robust and distributed manner, which means that concept formation is resilient to the removal of even seemingly important individual examples.[615]

332. Researchers have reached similar conclusions when studying data influence in the context of LLMs. For example, researchers have developed a scalable method for computing influence functions, which estimate how much each training sequence affects a model's output.[616] Applying this method to models ranging from 810 million to 52 billion parameters, they traced which training examples most contributed to specific model

---

completely different, many of which are public domain, that Claude would have no problem understanding what song lyrics are and how to write new ones without access to any of the lyrics indicated in this case, let alone outside of the case."); Declaration of Jared Kaplan in Support of Defendant's Opposition to Plaintiffs' Renewed Motion for Preliminary Injunction ("Kaplan Declaration") (Aug. 22, 2024), Dkt. 212-5, 291:6-292:8 ("As I indicated earlier, I think it's likely that music reviews, discussions of songs, discussions of music, of poetry are likely to be more relevant to Claude's behavior than song lyrics."); Deposition of Daniela Amodei 30(b)(1) (Anthropic) Transcript (Nov. 12, 2025), 180:7-180:11 ("It's possible that it, you know, learned that information from seeing song lyrics that were publicly available or from reading books about how people write songs or from reading other publicly available or purchased information.").

[614] Covert, Ian, Wenlong Ji, Tatsunori Hashimoto, and James Zou. "Scaling Laws for the Value of Individual Data Points in Machine Learning." arXiv:2405.20456. Preprint, arXiv, May 30, 2024. https://doi.org/10.48550/arXiv.2405.20456 ("we find that a data point's contribution to model's performance shrinks predictably with the size of the dataset in a log-linear manner.").

[615] Researchers removed the 10,000 training images most linked to a given concept and retrained the model. Neither the network layer where the concept emerged nor the sparsity of the concept representation changed significantly. This finding suggests that neural networks learn concepts in a robust and distributed manner, rather than relying on a few training examples. This also suggests that concept formation in neural networks is resilient to the removal of seemingly important individual examples. *See,* Konz, Nicholas, Charles Godfrey, Madelyn Shapiro, Jonathan Tu, Henry Kvinge, and Davis Brown. "Attributing Learned Concepts in Neural Networks to Training Data." arXiv:2310.03149. Preprint, arXiv, December 28, 2023. https://doi.org/10.48550/arXiv.2310.03149.

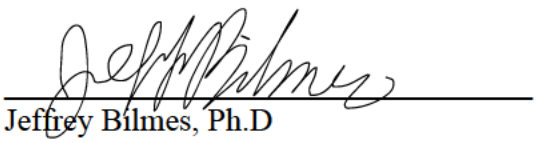
[616] Grosse, Roger, Juhan Bae, Cem Anil, et al. "Studying Large Language Model Generalization with Influence Functions." arXiv:2308.03296. Preprint, arXiv, August 7, 2023. https://doi.org/10.48550/arXiv.2308.03296.

ERRATA TO THE OPENING EXPERT REPORT OF JEFFREY BILMES, PH.D

I hereby make the following changes to correct citation errors appearing in the Expert Report of Jeffrey Bilmes, Ph.D, dated December 19, 2025:

FOOTNOTE	MATERIALS CONSIDERED ENTRY	CHANGE	EXPLANATION OF CHANGE
30	340	"Clio: Privacy-Preserving Insights" to "Clio: A System for Privacy Preserving Insights"	Amend title to cite to the title used on Anthropic's webpage instead of the title of the underlying Clio study contained within that webpage.
102		"'Scaling Laws and Interpretability of Learning from Repeated Data.' arXiv:2205.10487. Preprint, arXiv, May 21, 2022. https://doi.org/10.48550/arXiv.2205.10487. (Anthropic_0000520123)." to "'Studying Data Repetition Through Scaling Laws and Mechanistic Interpretability,' (draft) [Anthropic_0000520123]."	Amend citation and title to use the title of an earlier draft of study found at Anthropic_0000520123 instead of the title in the published version of the Scaling Laws paper.
188	284	"Anthropic's Am. Supp. Resp. & Obj. to Interrogatory 9 (Oct. 9, 2025)." to "Anthropic's Am. Supp. Resp. & Obj. to Interrogatory 6 (Oct. 9, 2025)."	Correct citation after inadvertently citing to wrong interrogatory number and omitting interrogatory number from materials considered list.
557, 558, 559	289	"Anthropic's Supp Resp. & Obj. to Interrogatory 18" to "Anthropic's Am. Supp. Resp. & Obj. to Interrogatory 18 (Nov. 5, 2025)."	Correct citation format.
475	159	"Bart Harbison" to "Bart Herbison"	Correct spelling error.

Signed in Seattle, Washington, on January 17, 2026.



Jeffrey Bilmes, Ph.D

SECOND ERRATA TO THE OPENING EXPERT REPORT OF JEFFREY BILMES, PH.D

I hereby make the following changes to correct inadvertent omissions appearing in the Expert Report of Jeffrey Bilmes, Ph.D, dated December 19, 2025:

PAGE/ PARAGRAPH	CHANGE	EXPLANATION OF CHANGE
Page 5, Paragraph 3	Append to end, "I am being compensated at my hourly rate of $1,000 for my work in this matter. Neither my compensation, nor that of my support team (Keystone), is affected by the outcome of this matter. In addition, I have not testified at deposition or trial in the last four years."	Inadvertently omitted

Signed in Seattle, Washington, on February 26, 2026.



Jeffrey Bilmes, Ph.D