EXHIBIT 4

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL., Plaintiffs, v. ANTHROPIC PBC, Defendant.	Case Number: 5:24-cv-03811-EKL-SVK Judge Eumi K. Lee Magistrate Judge Susan van Keulen

EXPERT REPORT OF DR. BEN ZHAO

December 19, 2025

Contains Information Designated as
Confidential and/or Highly Confidential—Attorneys' Eyes Only

repetitive inclusion of certain such data in training corpora during the pre-training process may reinforce regurgitation of texts that developers such as Anthropic do not desire.[47]

62. Anthropic has also disclosed that it relies heavily on datasets other third parties have scraped from the internet (among other methods of collection), including the Common Crawl, The Pile, and OpenWebText2 datasets that include large amounts of copyrighted and unlicensed material, including Publishers' lyrics.[48] In particular, the Common Crawl includes unlicensed copies of Publishers' lyrics scraped from licensed lyrics websites like LyricFind.com and Musixmatch.com.[49] Meanwhile, The Pile is known to include the "YouTube Subtitles" sub-dataset, which contains various unlicensed copies of Publishers' lyrics compiled from the subtitles of licensed YouTube videos for those works.[50] Anthropic has also trained on data specifically taken from Reddit,[51] among various other datasets.

63. After obtaining its datasets, Anthropic trains Claude to its specifications, selecting the level of compression and data loss, the weights it will give different forms of data, and other criteria. The resulting "pre-trained" model can be described as a "text completion model[] in which you give an input string and then it . . . generates the next parts of that string."[52] This pre-trained model may be suited to various tasks but memorizes large amounts of its datasets; Anthropic's

---

[47] Katherine Lee et al., *Deduplicating Training Data Makes Language Models Better*, arXiv:2107.06499 (Mar. 24, 2022), https://arxiv.org/pdf/2107.06499.
[48] *A General Language Assistant as a Laboratory for Alignment* 8, 27, ANTHROPIC (Dec. 9, 2021), https://arxiv.org/pdf/2112.00861.pdf; Kaplan Decl. ¶¶ 24–25, ECF No 67-1; Anthropic's Amended and Supplemental Response to Plaintiffs' Interrogatory No. 6, dated Oct. 9, 2025, at 3.
[49] *See* Kevin Schaul et al., *Inside the secret list of websites that make AI like ChatGPT sound smart*, WASH. POST (Apr. 19, 2024), https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/.
[50] *See* Annie Gilbertson & Alex Reisner, *Apple, Nvidia, Anthropic Used Thousands of Swiped YouTube Videos to Train AI*, Proof News (Jul. 16, 2024), https://www.proofnews.org/apple-nvidia-anthropic-used-thousands-of-swiped-youtube-videos-to-train-ai/; Alex Reisner, *Search the YouTube Videos Secretly Powering Generative AI*, Proof News (Jul. 16, 2024), https://www.proofnews.org/youtube-ai-search/.
[51] Anthropic's Amended and Supplemental Response to Plaintiffs' Interrogatory No. 6, dated Oct. 9, 2025, at 3.
[52] *Id.* at 93:24–94:7.

internal documents concede that "Large LMs memorize a LOT."[53] Indeed, LLMs are probability models and are, fundamentally, memorizing how often certain words or phrases appear next to each other (encoded as conditional probabilities in the model). This leads to literal memorization and regurgitation of texts, as I have explained above.[54] In 2023, one Anthropic employee discovered that Anthropic's models were "memorizing in this harmful, superficial way . . . large amounts of text" and that "[t]he effect was getting worse at scale. It seemed like the bigger models were memorizing more[.]"[55] This occurs when a model is "trained on a piece of text many times."[56] Note that this is entirely consistent with the concept of LLM training as lossy compression, as I have also explained above.[57] A larger capacity model is more capable of compressing training data in its entirety, allowing more accurate memorization. Similarly, training on the same text multiple times teaches the lossy compression process to prioritize this text above others, increasing the likelihood of it being memorized by the model. It also helps the training process to find better compressed representations of the training data, which manifests itself in training as lower loss values in the optimization process. This is a feature and not a bug of LLMs and their training process, and something that developers might choose to mitigate due to degradation of LLM performance but cannot eliminate.[58] As a result, even if Anthropic were to apply aggressive data cleaning or input filtering, Claude would nonetheless be prone to memorization and regurgitation of texts that appear frequently in its training corpus.

---

[53] *See* D. Drain Dep. Ex. 140 (Anthropic_0000507796); *see also, e.g.*, Anthropic_0000458427 (July 2021 memo discussing whether "we can reduce data recitation / solve the copyright problem internal to the model via RL").
[54] *Supra,* Section IV. A. 2.
[55] D. Drain Dep. Tr. 90:2–24. *See also* Anthropic_0000500599-500647.
[56] D. Drain Dep. Tr. 71:4-7.
[57] *Supra*, Section IV.A.3.
[58] *See* Alexander Xiong, et al., *The Landscape of Memorization in LLMs: Mechanics, Measurement, and Migration* (July 8, 2025), arXiv:2507.05578, https://arxiv.org/pdf/2507.05578; Ali Satavy, et al., *Undesirable Memorization in Large Language Models: A Survey*, arXiv:2410.02650v1 (Oct. 3, 2024), https://arxiv.org/pdf/2410.02650v1; Stella Biderman, et al., *Emergent and Predictable Memorization in Large Language Models*, arXiv:2304.11158 (Apr. 21, 2023), https://arxiv.org/pdf/2304.11158.

Dated: December 19, 2025

Ben Zhao