OPPENHEIM + ZEBRAK, LLP
Matthew J. Oppenheim
Nicholas C. Hailey
Audrey L. Adu-Appiah
(admitted *pro hac vice*)
4530 Wisconsin Ave., NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
matt@oandzlaw.com
nick@oandzlaw.com
aadu-appiah@oandzlaw.com

Jennifer Pariser
Andrew Guerra
Timothy Chung
(admitted *pro hac vice*)
461 5th Avenue, 19th Floor
New York, NY 10017
Telephone: (212) 951-1156
jpariser@oandzlaw.com
andrew@oandzlaw.com
tchung@oandzlaw.com

COBLENTZ PATCH DUFFY & BASS LLP
Jeffrey G. Knowles (SBN 129754)
One Montgomery Street, Suite 3000
San Francisco, CA 94104
Telephone: (415) 391-4800
ef-jgk@cpdb.com

COWAN, LIEBOWITZ & LATMAN, P.C.
Richard S. Mandel
Jonathan Z. King
Richard Dannay
(admitted *pro hac vice*)
114 West 47th Street
New York, NY 10036-1525
Telephone: (212) 790-9200
rsm@cll.com
jzk@cll.com
rxd@cll.com

Attorneys for Plaintiffs

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 **REPLY DECLARATION OF MICHAEL D. SMITH IN SUPPORT OF PLAINTIFFS' REPLY IN SUPPORT OF MOTION FOR PRELIMINARY INJUNCTION** REDACTED Hon. Eumi K. Lee

I, Michael D. Smith, hereby declare pursuant to 28 U.S.C. § 1746:

1. I am the J. Erik Jonsson Professor of Information Technology and Marketing at Carnegie Mellon University's Heinz College of Information Systems Management and Public Policy Management.

2. I submit this Reply Declaration in connection with Plaintiffs' Reply in Support of the Motion for Preliminary Injunction filed by Concord Music Group, Inc., Capitol CMG, Inc., Universal Music Corp., Songs of Universal, Inc., Universal Music – MGB NA LLC, Polygram Publishing, Inc., Universal Music – Z Tunes LLC, and ABKCO Music, Inc. (collectively, "Publishers" or "Plaintiffs"), in their lawsuit against Defendant Anthropic PBC ("Anthropic").

3. My statements set forth below are based on my specialized knowledge, education, and experience, as applied to the facts and circumstances of this case. If called upon, I would and could competently testify as to the matters contained herein.

4. This Declaration incorporates by reference my previous declaration, ECF No. 182 ("August 1 Declaration"), submitted in support of Plaintiffs' Renewed Motion for Preliminary Injunction, ECF No. 179 ("Mot."). My previous declaration contains important background and details not repeated here, including my curriculum vitae, areas of expertise, and initial conclusions relating to (1) the economic impact of Anthropic's use of Publishers' copyrighted works on Plaintiffs and the songwriters they represent; and (2) the economic feasibility of complying with the preliminary injunction Publishers seek.

5. In preparing this Ddeclaration, I relied on my general background and training and reviewed Publishers' filings, my previous declaration, Anthropic's Opposition to Plaintiffs' Motion for Preliminary Injunction, ECF No. 207 ("Anthropic's Opposition" or "Opp."), supporting filings, and other documents referenced herein.

6. The purpose and intent of this declaration is to provide a limited response to assertions made in Anthropic's Opposition and the supporting Declarations of Dr. Jared Kaplan, ECF No. 209 ("Kaplan Decl."); Dr. Steven Peterson, ECF No. 210 ("Peterson Decl."); and Ms. Dawn Hall, ECF No. 211 ("Hall Decl.").

Licensing Market for Copyrighted Content

7. Anthropic argues that "[i]t would not be possible for a generative AI platform like Anthropic to amass sufficient content to train an LLM like Claude in arm's length licensing transaction**, *at any price***." Opp. 6 (emphasis added). Dr. Peterson also suggests that a market for

licensing lyrics could not exist because the value of licensing copyrighted content for use by AI companies would not overcome transaction costs. Peterson Decl. ¶¶ 33–37.

8. Plainly, these arguments are disproven by the myriad license agreements entered by OpenAI and other AI companies for relatively small libraries of copyrighted content. *See, e.g.,* August 1 Declaration ¶¶ 76–77.

9. For example, consider the May 2024 agreement between OpenAI and *The Wall Street Journal*. Based on my preliminary analysis of works available on ProQuest's TDM Studio, *The Wall Street Journal* content published between the years 1923 and mid-2024 constitutes roughly 1.6 billion words. This number is remarkably small even relative to the "trillions of tokens," as Anthropic posits, "that general-purpose LLMs require." Opp. 6. In other words, the size of OpenAI's total training corpus likely dwarfs the number of tokens gained from *The Wall Street Journal*'s content. Applying Anthropic's logic, *The Wall Street Journal*'s content would therefore only have a small marginal value to OpenAI and likely not enough to overcome any necessary transaction costs. *See, e.g.,* Peterson Decl. ¶¶ 18–20; 24–25, 33–37. However, the net value of the license for Wall Street Journal content ***exceeded $250 million***. Alexandra Bruell et al., *OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over $250 Million*, WALL STREET JOURNAL (May 22, 2024), https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba. This $250 million price tag is no small sum and demonstrates even relatively small libraries of content can add sufficient value to LLM developers to warrant large payments. Anthropic's suggestion that there is or can be no market for AI companies to negotiate with rightsholders because there is no clearing price from such negotiations is demonstrably false.

10. Dr. Peterson also cites OpenAI's failed negotiations with *The New York Times* in support of the notion that the *Times* lacked "sufficient data to make it worthwhile for OpenAI to expend the cost and effort to negotiate a license to train." Peterson Decl. ¶ 34.

11. Dr. Peterson surmises that a 100-year archive *New York Times* would likely yield 5.5 billion words and offered little marginal value to OpenAI. Peterson Decl. ¶ 25. If that were true, then the 1.6 billion works in *The Wall Street Journal* archive would not have come with a $250 million price tag to OpenAI. Additionally, Dr. Peterson's line of reasoning arrives at the

incorrect conclusion. That OpenAI spent time and resources to participate in discussions presupposes that *The New York Times*' content was valuable enough to OpenAI to enter discussions in the first place. Moreover, Dr. Peterson's conclusion about why the negotiations failed is nothing more than speculation.

12. Anthropic also suggests that licensing markets for copyrighted lyrics are materially different from licensing markets for "news, magazine, social media, images, and audio rights" and therefore evidence of agreements for the latter do not establish a "*bona fide market*" for Publishers' copyrighted lyrics. Opp. 23-24.

13. This argument artificially narrows the market relevant to the analysis and fails to address or raise the salient economic question, which is whether ***the economic characteristics*** of functioning licensing markets for ***images, news, or social media content*** are materially different from ***the economic characteristics*** for a functioning licensing market for ***lyrics***. Lyrics may be different from newspapers and magazines in certain ways, but those differences are largely irrelevant to the analysis at hand.

14. The relevant analysis is whether the transactions (licensing agreements) for the goods exchanged or services rendered (copyrighted lyrics or other copyrighted materials) lead to welfare-enhancing outcomes. That is, if the value of the licensed content, net transaction costs, is positive, then a welfare-enhancing agreement is made. In this regard, the economic characteristics necessary to support markets for lyrics and other forms of licensed content, such as "news, magazine, social media, images, and audio," are essentially identical, if not coinciding. Anthropic offers no compelling reason to believe the economic fundamentals of licensing markets for the lyrics represented in the musical compositions owned by Publishers is any different.

License Agreements Contemplate Training and Other Uses for Copyrighted Works

15. Anthropic and Dr. Peterson assert the agreements cited by Publishers, Mot. 22, and referenced in my previous declaration, August 1 Declaration ¶¶ 76–77, do not provide a basis for the Court to find a developing market for licensing lyrics for AI training purposes. *See* Opp. 22; Peterson Decl. ¶¶ 38–39. In doing so, Anthropic relies on the following arguments.

16. ***First***, Anthropic's argument that since the "agreements are not in the record," "it [is] impossible to determine whether they in fact constitute bare licenses to train generative AI models," Opp. 23, ignores ample evidence pointing to the contrary.

17. While it may come as no surprise that I, like other members of the public, do not have access to copies of confidential agreements between AI companies and rightsholders, the suggestion that the inability to review the agreements themselves tends to negate the existence of a developing market for content licensing market for training AI is infirm. Opp. 23; Peterson Decl. ¶ 39. This ignores the robust press coverage surrounding these agreements has described these deals for copyrighted content as permitting AI training, and not simply for the convenience of accessing content that otherwise could have been used under the fair use doctrine. For example, *The Wall Street Journal*'s ***own coverage*** of News Corp's deal with OpenAI for *Journal* content referenced "people familiar with the situation" and described the deal as "a major content-licensing pact" allowing OpenAI to use *Journal* content to "train its technology." Alexandra Bruell et al., *OpenAI, WSJ Owner News Corp Strike Content Deal Valued at Over $250 Million*, WALL STREET JOURNAL (May 22, 2024), https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba.

18. Other examples abound. *See, e.g.,* Matt O'Brien, *ChatGPT-maker OpenAI signs deal with AP to license news stories*, AP NEWS (Jul. 13, 2023), https://apnews.com/article/openai-chatgpt-associated-press-apf86f84c5bcc2f3b98074b38521f5f75a ("The AP deal is valuable to a company like OpenAI because it provides a trove of material that it can use for training purposes"); Sara Fischer, *Exclusive: Time strikes licensing deal with OpenAI*, AXIOS (Jun. 27, 2024), https://www.axios.com/2024/06/27/openai-time-licensing-deal-chatgpt ("The deal gives OpenAI access to Time's archives from the last 101 years to train its large language models"); Sara Fischer, *OpenAI inks licensing deal with Dotdash Meredith,* AXIOS (May 7, 2024), https://www.axios.com/2024/05/07/openai-dotdash-meredith-licensing-deal ("Dotdash Meredith . . . inked a deal with OpenAI to license its content to train OpenAI's algorithms"); Louis Dreyfus and Jérôme Fenoglio, *Le Monde and Open AI sign partnership agreement on artificial intelligence,* LE MONDE (Mar. 13, 2024),

https://www.lemonde.fr/en/aboutus/article/2024/03/13/le-monde-signs-artificial-intelligence-partnership-agreement-with-openai_6615418_115.html ("[the deal] covers both the training of artificial intelligence models developed by the American company and answer engine services such as ChatGPT").

19. Academic publisher Taylor & Francis' recent AI Partnership and Data Access Agreement with Microsoft, which grants Microsoft "access to . . . content and data to help improve relevance and performance of AI systems," *Informa PLC: Market Update*, INFORMA (May 8, 2024) at 3, https://www.informa.com/globalassets/documents/investor-relations/2024/informa-plc---market-update.pdf, is also alleged to provide "access to [Taylor & Francis authors'] content for [AI] training [purposes]," *see* ColonelCrikey, REDDIT (Aug. 2024), https://www.reddit.com/r/publishing/comments/1efy2nu/i_emailed_my_editor_at_taylor_francis_to_find_out/.

20. Stakeholders and parties to these agreements have also clearly indicated that AI training is covered by these deals. For instance, in announcing its licensing deal with Axel Springer, which owns media brands including *Business Insider, Politico,* and European publications *BILD* and *WELT*, *see* August 1 Declaration ¶ 76, OpenAI stated the "use of quality content from Axel Springer media brands" also included "the ***training of OpenAI's sophisticated language models***." *Partnership with Axel Springer to deepen beneficial use of AI in journalism*, OPENAI (Dec. 13, 2023), https://openai.com/index/axel-springer-partnership/ (emphasis added).

21. I also note that in the time since my last declaration, OpenAI entered into a major new licensing agreement with Condé Nast, owner of *The New Yorker, Vogue*, and *Wired* magazines, which was reported to include "permission to train on Condé Nast content," according to a Condé Nast spokesperson. Kyle Wiggers, *OpenAI signs deal to train on Condé Nast content, surface stories in ChatGPT*, TECHCRUNCH (Aug. 20, 2024), https://techcrunch.com/2024/08/20/openai-signs-deal-to-train-on-conde-nast-content-surface-stories-in-chatgpt/.

22. ***Second***, Dr. Peterson claims the terms of the agreements, including those referenced above, "are not clear and appear to encompass more than the licensing of training data." Peterson Decl. ¶ 39. However, the fact that some licensing agreements between AI companies like

OpenAI and rightsholders might encompass uses beyond training does not tend to make the existence of a licensing market for training purposes less likely. Nor does it tend to show that the licensing market for Publishers' lyrics are unaffected by Anthropic's conduct.

23. Dr. Peterson cites the agreement between OpenAI and Axel Springer as an example of an agreement "with copyright owners supplying data to AI firms to allow for the output of the AI firm's models to ***reproduce*** the training data to a degree that is beyond fair use," Peterson Decl. ¶ 43 (emphasis added), including by allowing ChatGPT to "show parts of articles in [its] responses," *id.* (quoting Will Oremus and Elahe Izadi, *AI's future could hinge on one thorny legal question*, THE WASHINGTON POST (Jan. 4, 2024), https://www.washingtonpost.com/technology/2024/01/04/nyt-ai-copyright-lawsuit-fair-use/). Dr. Peterson also suggests "a license for access to data, or a broader deal allowing the display of copyrighted material . . . may be less expensive than defending a lawsuit." *Id.* ¶ 44.

24. If anything, Dr. Peterson demonstrates that the licensing market for copyrighted content by AI companies expands beyond training uses to include uses that other licensees of Publishers' lyrics typically contract for, such as the reproduction, distribution, and display of copyrighted content. *See, e.g.,* Declaration of David Kokakis, ECF No. 187 ¶¶ 10–12; Declaration of Kent Draughon, ECF No. 186 ¶¶ 10–12; Declaration of Alisa Coleman, ECF No. 185 ¶¶ 10–12; Declaration of Duff Berschback, ECF No. 184 ¶¶ 10-12. In fact, unauthorized reproductions, distributions, and displays of copyrighted content form the basis of some of Publishers' legal claims against Anthropic. *See* Complaint, ECF No. 1 ¶¶ 112, 120, 134. Thus, Dr. Peterson establishes that a market exists for Publishers' lyrics for uses by AI companies *in addition* to training to include uses that unambiguously and directly complete with Publishers' existing licensees.

25. ***Third***, Anthropic argues that OpenAI's public position in favor of fair use "casts doubt on [my] assumption that these deals must be about licensing data for LLM training." Opp. 23. Dr. Peterson suggests OpenAI and other AI companies may "license data for purposes that are consistent with their views that training LLMs with copyrighted data is fair use" and that "[t]he

existence of a handful of licenses between OpenAI and content providers does not support the inference that OpenAI has come to the conclusion that it must license data." Peterson Decl. ¶ 41.

26. I draw precisely the opposite conclusion from OpenAI's decision to pay millions of dollars to license content on time-limited bases, August 1 Declaration ¶ 76: if OpenAI was confident in the fair use defense, there would be no need to sign these deals in the first place. OpenAI's licensing decisions are costly, *id.* ¶ 78, and OpenAI's decision to enter into licenses for content that it could likely obtain through the internet speaks volumes, despite OpenAI's public position on fair use.

27. ***Finally***, Dr. Peterson's assertion that "the handful of agreements that appear to cover rights much broader than training do not support the economic conclusion that AI firms are moving toward licensing, or could license, ***all*** of their training data," Peterson Decl. ¶ 11 (emphasis added), assumes that Publishers require all AI companies to seek authorization for all the content used in their training corpora. This is not the question being asked. The proper question is whether there is a market for licensing or seeking authorization for ***copyrighted*** data in training corpora. The answer is demonstrably yes. Additionally, the argument that it would be economically impractical for LLMs to negotiate licenses or seek authorization from *all* rightsholders and therefore it is economically impractical for LLMs to negotiate licenses with *any* rightsholders, *see* Peterson Decl. ¶ 45, not only suffers from the fallacy of composition—what is true for one member of the group does not make the same true for the entire group—but it also suffers from the empirical problem that plenty of rightsholders *have* negotiated multi-million-dollar licensing agreements with LLMs.

Economic Feasibility of Publishers' Injunctive Relief

28. Ms. Hall argues that the removal of Publishers' lyrics from Anthropic's training corpus would [redacted] Hall Decl. ¶ 68. However, Anthropic co-founder and Chief Science Officer Jared Kaplan admits [redacted] and [redacted] is a necessary step in developing Anthropic's newer and subsequent models. Kaplan Decl., ¶ 60. In their request for preliminary injunction, Publishers have asked only that

Anthropic remove their copyrighted content from future versions of Claude for which training has not yet begun. *See* Mot. 30; *Id.* Ex. A, ECF No. 179-1. [redacted]

29. Ms. Hall also argues that any damages that might occur because of Anthropic's use of lyrics are easily quantified through "standard economic models" because "[p]laintiffs routinely enter into license agreements for rights to the Works and recognize that compensation for use of their Works can be quantified." Hall Decl. ¶¶ 14, 34. However, that statement does not consider that there are unique challenges around quantifying the harm in this case, including, for example, the harm to Publishers' relationships with their songwriters, the denial of proper attribution and goodwill, and reputational harms, which cannot be easily, if ever, distilled into a single monetary value.

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct to the best of my personal knowledge and belief.

Michael D. Smith

Executed in Pittsburgh, PA, this 12th day of September, 2024.