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 ED NEWTON-REX IN SUPPORT OF PLAINTIFFS' REPLY IN SUPPORT OF MOTION FOR PRELIMINARY INJUNCTION** Hon. Eumi K. Lee

I, Ed Newton-Rex, hereby declare pursuant to 28 U.S.C. § 1746:

1. I am the CEO and founder of Fairly Trained, a non-profit organization that certifies generative AI companies that obtain licenses for the data they use or use public domain or openly licensed data.

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"), 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 dated August 1, 2024, ECF No. 183 ("August 1 Declaration"), submitted in support of Plaintiffs' Motion for Preliminary Injunction, ECF No. 179. My earlier declaration contains important background and details not repeated here, including my curriculum vitae, areas of expertise, and initial conclusions relating to the Anthropic's copying of Publishers' works in connection with Anthropic's Claude models and the market for Publishers' works.

5. In preparing this Declaration, I relied on my general background and training and I reviewed Publishers' filings, my previous declaration, Anthropic's Opposition to Plaintiffs' Motion for Preliminary Injunction ("Anthropic's Opposition" or "Opp.") and supporting filings, and other documents referenced herein.

Market for Licensed Content in AI Models

6. As described in my prior Declaration and that of Professor Smith dated August 1, 2024, ECF 182, the market for licensed content in AI models is evolving rapidly. New deals are reported regularly, demonstrating the interest of responsible AI companies in training their AI models on licensed and authorized content. The AI music market has been particularly active in launching new businesses based on licensed content. In fact, my organization Fairly Trained has granted certifications to a dozen companies since the beginning of 2024 in the AI music space. These include Rightsify, whose AI music model Hydra II was trained on over one million songs, using only licensed content, Stuart Dredge, *Rightsify unveils 'most diverse and truly ethical' AI music model*, MUSICALLY (Mar. 21, 2024), https://musically.com/2024/03/21/rightsify-unveils-

most-diverse-and-truly-ethical-ai-music-model/; and Jen, whose AI music model was trained on over 40 fully-licensed catalogs, Kristin Robinson, *Futureverse Launches Jen, An Ai Music Model Focused on 'Transparency*,'" BILLBOARD (June 20, 2024), https://www.billboard.com/pro/futureverse-jen-ai-music-model-launch/. Fifty companies have signed the "aiformusic" principles, pledging to respect copyright in AI music generation. Julian Schmauch, *AIformusic: 50 tech companies sign new guidelines for AI in music*," Gearnews (June 25, 2024), https://www.gearnews.com/aiformusic-50-tech-companies/#the-big-50-these-companies-have-signed-a-iformusics-agreement.

7. This developing interest in licensed content is not limited to the area of AI music. Just last month, for example, OpenAI and Conde Nast announced a deal in which OpenAI's products, such as ChatGPT, will be licensed to display content from *Vogue*, *Architectural Digest*, *Vanity Fair*, *Wired*, and other Conde Nast properties. Hayden Field, *OpenAI announces content deal with Conde Nast, including content from Wired, The New Yorker and Vogue,* CNBC, August 20, 2024. https://www.cnbc.com/2024/08/20/openai-announces-content-deal-with-cond-nast-including-content-from-wired-the-new-yorker-and-vogue.html.

8. In his Declaration, Dr. Peterson argues that developers of large language models cannot practically or affordably license all the content they want to train their models and therefore, there is no market for any of it. Declaration of Steven Peterson dated August 22, 2024, ECF 210 ("Peterson Decl.") ¶ 37. He supports this conclusion in part by citing the large amount of data required to train these models. But the data used to train Claude includes synthetic data—that is, data created by an LLM. Declaration of Jared Kaplan dated August 22, 2024, ECF 209 ("Kaplan Decl.") ¶ 28 ("Anthropic may … includ[e] … synthetic data generated using our existing LLMs"); *see also* Declaration of Ben Y. Zhao, ¶ 9. Indeed, using synthetic data as part of an LLM's training data is commonplace: the CEO of LLM developer Cohere has said that "[Synthetic data] is already huge . . . even if it's not broadcast widely." Madhumita Murgia, *Why computer-made data is being used to train AI Models*, FINANCIAL TIMES (July 18, 2023), https://www.ft.com/content/053ee253-820e-453a-a1d5-0f24985258de. Dario Amodei, CEO of Anthropic, has even said, "I think it may be possible to get an infinite data generation engine" when discussing the use of synthetic data.

CNBC, *Anthropic CEO Dario Amodei on Claude 3 model, AI arms race and Big Tech partnerships,* YOUTUBE (Apr. 23, 2024), https://www.youtube.com/watch?v=Y2N-uhZFPIM. So, LLM providers do not need to compile all of their training data from third parties: some of it they can, and do, generate themselves. Anthropic has not revealed how much of Claude's training data is synthetic data; without knowing this, it is impossible to say how difficult it would be to license enough data to provide the copyrighted elements of the dataset required to train a high-performing LLM.

9. In my view, Anthropic's itemized impediments to licensing or seeking authorization from copyright owners in building high-performing LLMs are exaggerated. This is for several reasons. First, as discussed above, the size of Claude's training set is an overestimate of the amount of content Anthropic would need to license, since by their own admission the training set includes synthetic data, and, as discussed in my previous declaration, there are also large public domain datasets, such as Common Corpus, which can be combined with licensed data, ECF No. 183 ¶ 20. Second, the increasing number of agreements for licensed content in the AI space, at least some of which are known to cover training, suggests it is not impossible to find license rates and transaction costs that are amenable to both AI companies and rights holders. And third, efforts by LLM developers to date have overwhelmingly relied at least in part on web-scraped, presumably often copyrighted data, leaving the question of whether or not a high-performing LLM can be trained by combining public domain data, licensed data, and synthetic data—and avoiding web-scraped, copyrighted data—largely untested. In these circumstances, it seems premature to conclude that doing so would be impossible.

10. In his Declaration, Jared Kaplan suggests that LLMs are always better the more content is available to them. Kaplan Decl. ¶ 24. If we accept that as true, it would also be true for many products that require copyrighted content. A music streaming service is a better product if users can use it to access any music they like. This does not in itself justify removing the requirement that uses of content be authorized.

11. Of course, it takes effort and expense to negotiate licensing deals, and the market for licensed content will take time to evolve, but in my view, the evidence shows a trend towards AI companies licensing content.

Claude's Guardrails

12. Anthropic contends its guardrails are effective and copyrighted lyrics can no longer be obtained from Claude. Opp. 8. My review of public comments and my own testing suggests this is not true. For example, on June 29, 2024, an X user who calls themselves Pliny the Liberator (with the handle @elder_plinius) posted a simple prompt that he claimed was effective in obtaining otherwise blocked outputs from Claude. *See* **Exhibit A**. Pliny is well-known as a source of methods to 'jailbreak' AI models, and this post has been available since June 29, 2024. Indeed, at time of writing it is the top hit on X for the search 'claude pliny'.

13. Using a version of this prompt, I asked Claude for the lyrics to 46 of the Publishers' works at issue in this lawsuit and was able to obtain a complete set or substantial portion of all of them. *See* **Exhibit B**. In particular, this included lyrics to "A Thousand Miles," "Ain't No Sunshine," "Because I Got High," "Brimful of Asha," "California Dreamin'," "Candle In The Wind," "Dog Days Are Over," "Every Breath You Take," "Fields of Gold," "Fresh Prince of Bel-Air," "God Only Knows," "Good Vibrations," "I Believe I Can Fly," "I Get Around," "I Miss You," "I Will Survive," "I Will Wait," "In My Place," "In The End," "Let's Get Together," "Little Lion Man," "Love Is All Around," "Milkshake," "Paradise," "Praise You," "Roxanne," "She Will Be Loved," "(Sittin' On) The Dock Of The Bay," "Somebody To Love," "Something Just Like This," "Somewhere Only We Know," "Songbird," "Sweet Caroline," "Sympathy for the Devil," "The Boys Are Back In Town," "The Scientist," "The World's Greatest," "These Boots Are Made For Walkin'," "Tiny Dancer," "To Make You Feel My Love," "Torn," "Violet Hill," "Viva La Vida," "Wouldn't It Be Nice," "Yellow," and "Your Song." In order to obtain these lyrics, I instructed Claude to insert a space between each letter of the output, which bypassed Claude's output filter and had no impact on the usability of the result. Using this method, I believe it would be a trivial matter to obtain the lyrics for any well-known song a user wanted.

14. I also prompted Claude to create "mashups" of several songs: "Brimful of Asha" with "Candle in the Wind"; "Because I Got High" with "Vida La Vida"; "Love Is All Around" with "Milkshake"; and "Fresh Prince of Bel Air" with "Fields of Gold." Again, I used a version of the prompt shared by Pliny the Liberator, asking Claude to insert a space between each letter of the output. *See* **Exhibit C**. Claude generated output copying portions of the lyrics to these works in response to each prompt, typically intermingling lyrics from multiple works.

15. Finally, I prompted Claude to generate outputs using Publishers' lyrics to promote certain products or endorse political candidates, again using the same method. I requested that Claude create a version of "Roar" to advertise Bud Light; "Vida La Vida" to advertise cigarettes; "You Can't Always Get What You Want" to endorse Donald Trump; and "How Soon Is Now" to endorse Donald Trump. *See* **Exhibit D**.

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.



Ed Newton-Rex

Executed in Palo Alto, CA, this 12th day of September, 2024.