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

PLAINTIFFS' THIRD SET OF REQUESTS FOR PRODUCTION OF DOCUMENTS AND THINGS TO DEFENDANT

Hon. Eumi K. Lee

CONTAINS INFORMATION DESIGNATED HIGHLY CONFIDENTIAL BY DEFENDANT ANTHROPIC PBC

Pursuant to Federal Rules of Civil Procedure 26 and 34, Plaintiffs 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. (each individually a "Publisher" and, collectively, "Plaintiffs" or "Publishers") serve their Third Set of Requests for Production of Documents and Things ("Requests") upon Defendant Anthropic PBC ("Defendant" or "Anthropic"). Plaintiffs request that Defendant respond in writing in accordance with the Federal Rules of Civil Procedure and the Definitions and Instructions herein and produce any documents or things at the offices of Oppenheim + Zebrak, LLP, 4530 Wisconsin Avenue NW, 5th Floor, Washington, DC 20016, or at such other time and place as the parties may agree upon, within thirty (30) days of the date of service.

DEFINITIONS

Unless otherwise indicated, the following definitions apply to these Requests:

1. The terms "Defendant," "Anthropic," "you," or "your" mean Anthropic PBC, including its parents, subsidiaries, predecessors, successors, affiliates, officers, directors, employees, contractors, freelancers, and/or any other person or entity currently or previously acting on behalf or at the direction of Anthropic PBC.

2. The terms "Plaintiffs" or "Publishers" refer collectively to Concord Music Group, Inc.; Capitol CMG, Inc. d/b/a Ariose Music, d/b/a Capitol CMG Genesis, d/b/a Capitol CMG Paragon, d/b/a Greg Nelson Music, d/b/a Jubilee Communications, Inc., d/b/a Meadowgreen Music Company, d/b/a Meaux Hits, d/b/a Meaux Mercy, d/b/a River Oaks Music, d/b/a Shepherd's Fold Music, d/b/a Sparrow Song, d/b/a Worship Together Music, d/b/a Worshiptogether.com Songs; Universal Music Corp. d/b/a Almo Music Corp., d/b/a Criterion Music Corp., d/b/a Granite Music Corp., d/b/a Irving Music, Inc., d/b/a Michael H. Goldsen, Inc., d/b/a Universal - Geffen Music, d/b/a Universal Music Works; Songs of Universal, Inc. d/b/a Universal - Geffen Again Music, d/b/a Universal Tunes; Universal Music - MGB NA LLC d/b/a Multisongs, d/b/a Universal Music - Careers, d/b/a Universal Music - MGB Songs; Polygram Publishing, Inc. d/b/a Universal - Polygram International Tunes, Inc., d/b/a Universal - Polygram International Publishing, Inc., d/b/a Universal - Songs of Polygram International, Inc.; Universal Music - Z Tunes LLC d/b/a

New Spring Publishing, d/b/a Universal Music - Brentwood Benson Publishing, d/b/a Universal Music - Brentwood Benson Songs, d/b/a Universal Music - Brentwood Benson Tunes, d/b/a Universal Music - Z Melodies, d/b/a Universal Music - Z Songs; and ABKCO Music, Inc.

3. "AI" means artificial intelligence.

4. "Claude" means any Anthropic AI models, technologies, and products that are "within the scope of relevance" or fall within the definition of Claude as set forth in Magistrate Judge van Keulen's November 26, 2024 Order, ECF No. 273.

5. "Communicate" or "communication" means, without limitation, any transmittal, conveyance, or exchange of a word, statement, fact, thing, idea, document, instruction, demand, question, or other information in any medium, whether by written, oral, or other means, including but not limited to electronic communications and electronic mail.

6. "Concern," "concerning," "evidence," "evidencing," "relating to," "relates to," "related to," "referring or relating to," "referring to," "regarding," and "refer or relate to" are construed broadly to mean relating to, referring to, describing, evidencing, or constituting or in any way relevant within the meaning of Fed. R. Civ. P. 26(b)(1).

7. "Date" means the exact day, month, and year, if so ascertainable or, if not, the best approximation (including relationship to seasons and/or other events).

8. "Document" is synonymous in meaning and equal in scope to the usage of the term in Fed. R. Civ. P. 34(a) and includes Electronically Stored Information ("ESI") existing in any medium from which information can be translated into reasonably usable form, including but not limited to email and attachments, word processing documents, spreadsheets, graphics, presentations, images, text files, databases, instant messages, transaction logs, audio and video files, voicemail, internet data, computer logs, text messages, and backup materials. Publishers interpret ESI to mean information that is stored electronically as files, documents, or other data on computers, servers, mobile devices, online repositories, disks, USB drives, tape or other real or virtualized devices or digital media. By way of illustration and without limitation, documents include at least the following: originals, drafts, and all non-identical copies. If a draft document has been prepared in several copies that are not identical, or if the original identical copies are no

longer identical due to subsequent notation, each non-identical document is a separate document. For the avoidance of doubt, any Documents or data shall include their associated metadata.

9. "Electronically stored information" or "ESI" is information that is stored electronically as files, documents, or other data on computers, servers, mobile devices, online repositories, disks, USB drives, tape, or other real or virtualized devices or digital media.

10. The terms "identify," "identity," and "identification" mean, as the context shall make appropriate:

a. With respect to a person, to the extent known, the person's full name, present or last known address, and when referring to a natural person, additionally, the present or last known place of employment;

b. With respect to a document, to the extent known, the type of document, subject matter, date of the document, author(s), addressee(s), and recipient(s);

c. With respect to a telephone conversation, to the extent known, the participants' full names, present or last known place of employment, present or last known address, the date of the conversation, and the subject matter discussed;

d. With respect to an oral conversation other than a telephone conversation, to the extent known, the participants' full names, present or last known place of employment, present or last known address, the date of the conversation, and the subject matter discussed;

e. With respect to a corporation, partnership, association, or other entity, to the extent known, its full name and present or last known address; and

f. With respect to any event or action, to the extent known, the dates or approximate dates thereof, a description of what occurred, and all identifiable persons involved.

11. "Person" and "individual" mean any natural person or any business, legal, or government entity, or association—including, but not limited to, any proprietorship, partnership, firm, company, or corporation.

12. "Thing" has the full meaning ascribed to it by the Federal Rules of Civil Procedure.

13. "Training Corpus" means, for a generative AI model, the data used to train that model.

14. Plaintiffs' use of terms, phrases, and definitions is for convenience and no term, phrase, or definition shall be construed as an admission by Publishers.

INSTRUCTIONS

1. The following shall apply to all Requests:

   a. The singular form of each word shall be construed to include and encompass its plural, and the plural of each word shall be construed to include and encompass its singular form;

   b. The conjunctions "and" as well as "or" shall be construed either conjunctively or disjunctively, whichever is more inclusive in context;

   c. "Each" and "every" shall be construed to include "all" and "any," and vice versa; and

   d. The use of a verb in any tense shall be construed to include the use of the verb in all other tenses.

2. These Requests are continuing to the extent permitted by Rule 26(e) of the Federal Rules of Civil Procedure. If you receive or otherwise become aware of information responsive to any Request after you have served your responses to these Requests, you must promptly supplement your responses to these Requests to provide such information.

3. Unless otherwise indicated, these Requests seek information from 2020 through the present. These Requests shall be deemed continuing so as to require a prompt and further supplemental response if you obtain additional responsive information at any time between the time for the initial response and the time of hearing or trial.

4. Each Request shall operate and be responded to independently and, unless otherwise indicated, no Request limits the scope of any other Request. This instruction does not require you to re-produce documents to the extent they are duplicative documents responsive to other Requests.

5. In responding to these Requests, you are required to furnish relevant and responsive information that is available to you or subject to your reasonable inquiry, including information in the possession, control, or custody of your present and former attorneys, accountants, advisors, representatives, agents, employees, or other persons directly or indirectly employed by you, and anyone else otherwise subject to your control. All Documents that respond, in whole or in part, to any portion of the Requests below shall be produced in their entirety, including all attachments and enclosures.

6. If there are no Documents or Things responsive to a particular Request following a reasonable inquiry, you should state so in writing.

7. If any Document or Communication covered by this set of Requests is withheld by reason of a claim of privilege (including work-product immunity), you shall furnish within a reasonable timeframe a list providing information on the nature of each withheld Document sufficient to allow Publishers to assess the claim of privilege, as required by Fed. R. Civ. P. 26(b)(5).

8. In the event you contend that any of these Requests are objectionable, in whole or in part, state with particularity each such objection and the bases therefore and respond to the remainder of the Request to the extent that you are not objecting to it.

9. If you object to any Request on the grounds of overbreadth, specifically state the manner in which you contend the Request is overly broad and respond to the Request as narrowed to conform to such objection.

10. If any Request is ambiguous or unclear to you, you are requested to contact the undersigned counsel as soon as possible so that the Request can be clarified to avoid unnecessary delays in discovery.

11. If you do not agree with any Definition of the terms provided herein, you are instructed to provide a reasonable alternative Definition for that term, consistent with industry custom and usage.

12. Each of the Definitions and Instructions contained herein shall be fully applicable to each Request notwithstanding that a Definition or Instructions above may, in whole or in part,

be reiterated in a particular Request and notwithstanding that a particular Request may incorporate supplemental Instructions or Definitions. If, however, you provided any alternate Definitions in accordance with these Instructions, indicate whether and how, any answer to any Request would differ if you relied on your alternate Definition in responding to that Request instead of the Definitions provided.

13. Each and every copy of a Document that is different from the original because of stamps, indications of the recipient(s), handwritten notes, marks, attachments, or any other reason, is a separate Document that must be produced.

14. Pursuant to Federal Rule of Civil Procedure 34(b), you are instructed to produce Documents as they are kept in the usual course of business or organized and labeled to correspond with the categories in this set of requests. Documents shall be produced in their original folders, binders, covers, or containers, or photocopies thereof. Such Documents shall be designated as responsive to the particular request as designated herein.

15. If a Document otherwise responsive to a request was, but no longer is, in your actual or constructive possession, custody, or control, please state whether the Document (a) is lost or destroyed; (b) has been transferred to another person or entity either voluntarily or involuntarily; or (c) has been otherwise disposed of; and in each instance, (d) explain the circumstance surrounding the disposition, including the date or approximate date of the disposition.

REQUESTS FOR PRODUCTION

REQUEST NO. 67: All Documents and Communications concerning the high-level design of Claude, including but not limited to its design objectives, functionality, primary components, training, and finetuning requirements.

REQUEST NO. 68: Documents sufficient to show all "diligence" Anthropic conducts on the data it uses to train Claude.[1]

REQUEST NO. 69: Documents sufficient to show all sources used to train Claude that feature song lyrics, including but not limited to public web pages crawled by Anthropic itself,

[1] *See, e.g.*, Anthropic, *The Claude 3 Model Family: Opus, Sonnet, Haiku*, at 3, https://assets.anthropic.com/m/61e7d27f8c8f5919/original/Claude-3-Model-Card.pdf.

public web pages included in third-party datasets or text repositories used by Anthropic in training such as the Common Crawl, and non-public data provided to Anthropic by data labeling services, paid contractors, or other third parties.

REQUEST NO. 70: All Documents and Communications supporting Anthropic's assertion that "[o]f the datasets Anthropic has used to train Claude as of the date of these Responses, it believes the following three data sources are more likely to contain song lyrics than other datasets . . . [1] Common Crawl [redacted]"[2]

REQUEST NO. 71: All Documents and Communications regarding Anthropic's use of "Common Crawl and [redacted]" to train Claude.[3]

REQUEST NO. 72: All Documents and Communications regarding Anthropic's use of "YouTube subtitles" to train Claude.[4]

REQUEST NO. 73: All Documents and Communications regarding Anthropic's use of "[redacted]" to train Claude.[5]

REQUEST NO. 74: All Documents and Communications regarding Anthropic's use of the dataset The Pile to train Claude.

REQUEST NO. 75: All Documents and Communications concerning copies of training data created during Anthropic's process of training Claude, including but not limited to during the processes of data cleaning, deduplication, and data processing (including but not limited to normalization and tokenization).

REQUEST NO. 76: All Documents and Communications concerning the creation of "scrambled" tokenized data and the use of such data for training Claude, including but not limited to records of which such data were parsed during the training process.[6]

---

[2] Anthropic's Am. Supp. Response to Interrogatory No. 6 (Dec. 20, 2024).

[3] *Id.*

[4] *Id.*

[5] *Id.*

[6] Anthropic's Am. Supp. Response to Interrogatory No. 9 (Dec. 20, 2024).

REQUEST NO. 77: All Documents and Communications concerning the quantitative criteria operative in the training process of Claude to adjust model parameters, including but not limited to criteria to glean language principles and the relationship between the words and phrases in the Training Corpus.

REQUEST NO. 78: All Documents and Communications regarding the importance, benefit, or impact of including song lyrics in the Training Corpus for Claude, including but not limited to those supporting Anthropic's statement that "song lyrics as a category are an important piece of training data for teaching Claude about the unique use of language in poetic and lyrical texts."[7]

REQUEST NO. 79: Documents sufficient to show the portion or fraction of Claude's training data constituted by song lyrics, including but not limited to those concerning Anthropic's statement that "song lyrics constitute a minuscule fraction of Claude's training data."[8]

REQUEST NO. 80: All Documents and Communications concerning the inclusion, removal, deletion, alteration, or modification of copyright information (including but not limited to title, author name, copyright owner name, or copyright notice) from works included in the Training Corpus of Claude.

REQUEST NO. 81: All Documents and Communications regarding Anthropic's efforts to investigate or identify whether materials subject to copyright protection, including but not limited to song lyrics, are included in the Training Corpus for Claude.

REQUEST NO. 82: All Documents and Communications regarding Anthropic's efforts to investigate or identify whether materials for which copyright information (including but not limited to title, author name, copyright owner name, or copyright notice) has been removed or altered are included in the Training Corpus for Claude.

REQUEST NO. 83: All Documents and Communications regarding Anthropic's lack of knowledge as to the contents of the datasets on which it trains Claude, including but not limited to

---

[7] Anthropic's Opp'n to Publishers' Motion for Preliminary Injunction at 22 (Aug. 22, 2024), ECF No. 207.

[8] *Id.* at 6.

those concerning Anthropic's statement that it "[redacted]"[9]

REQUEST NO. 84: All Documents and Communications regarding instances in which rightsholders or companies have "[redacted]" including but not limited to Anthropic's consideration of any such requests, discussions regarding removal of any such data from training datasets, and outcomes of any such requests.[10]

REQUEST NO. 85: All Documents and Communications regarding Anthropic's operation of its web crawling system for Claude, including but not limited to Anthropic's response or conduct with respect to "robots.txt instructions and other signals that website operators use to indicate whether they permit crawling of the content on their sites."[11]

REQUEST NO. 86: All Documents and Communications regarding the memorization or regurgitation of specific texts by Claude, including but not limited to the over-indexing of specific texts in Anthropic's Training Corpus and Anthropic's deduplication efforts in connection with the memorization or regurgitation of specific texts.

REQUEST NO. 87: All Documents and Communications relating to Anthropic's evaluation and analysis of Claude usage and usage patterns, including but not limited to Anthropic's "Claude insights and observations" ("Clio") platform.

REQUEST NO. 88: All Documents and Communications concerning whether "[t]ypical Anthropic users" "request song lyrics from Claude."[12]

REQUEST NO. 89: Documents sufficient to show the contents of Anthropic's [redacted]" including but not limited to the "[redacted]" included therein.[13]

REQUEST NO. 90: All Documents and Communications regarding the consideration

[9] Anthropic's Am. Supp. Response to Interrogatory No. 6 (Dec. 20, 2024).

[10] Anthropic's Am. Supp. Response to Interrogatory No. 9 (Dec. 20, 2024).

[11] Anthropic's Am. Response to Interrogatory No. 11 (Dec. 20, 2024).

[12] Anthropic's Opp'n to Publishers' Motion for Preliminary Injunction at 3 (Aug. 22, 2024), ECF No. 207.

[13] Anthropic's Am. Supp. Response to Interrogatory No. 12 (Dec. 20, 2024).

and determination of which song lyrics to include in Anthropic's [redacted].

REQUEST NO. 91: All Documents and Communications concerning the methodology used by Anthropic's Content Compare System to compare Claude output against the [redacted].

REQUEST NO. 92: All Documents and Communications regarding Anthropic's potential or actual indemnification of copyright infringement claims, including but not limited to the copyright indemnification provision of Anthropic's Commercial Terms of Service.

REQUEST NO. 93: All Documents on which Anthropic relies in connection with its fair use defense.

Dated: January 27, 2025

Respectfully submitted,

/s/ Nicholas C. Hailey

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

CERTIFICATE OF SERVICE

I hereby certify that on January 27, 2025, I caused the foregoing document to be served on the following counsel via email:

LATHAM & WATKINS LLP
Joseph R. Wetzel (SBN 238008)
joe.wetzel@lw.com
Andrew M. Gass (SBN 259694)
andrew.gass@lw.com
Brittany N. Lovejoy (SBN 286813)
brittany.lovejoy@lw.com
Ivana Dukanovic (SBN 312937)
ivana.dukanovic@lw.com
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: +1.415.391.0600

LATHAM & WATKINS LLP
Sarang V. Damle (*pro hac vice*)
sy.damle@lw.com
555 Eleventh Street NW, Suite 1000
Washington, D.C. 20004
Telephone: +1.202.637.2200

LATHAM & WATKINS LLP
Allison L. Stillman (*pro hac vice*)
alli.stillman@lw.com
1271 Avenue of the Americas
New York, New York 10020
Telephone: +1.212.906.1747

LATHAM & WATKINS LLP
Rachel Horn (SBN 335737)
rachel.horn@lw.com
140 Scott Drive
Menlo Park, California 94025
Telephone: +1.650.328.4600

/s/ Nicholas C. Hailey
Nicholas C. Hailey