# Exhibit 22

SONAL N. MEHTA (SBN 222086)
sonal.mehta@wilmerhale.com
ALLISON BINGXUE QUE (SBN 324044)
allison.que@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
2600 El Camino Real, Suite 400
Palo Alto, CA 94306
Telephone: (650) 858-6000

JOSEPH R. WETZEL (SBN 238008)
joe.wetzel@lw.com
**LATHAM & WATKINS LLP**
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: (415) 391-0600

*Attorneys for Defendant*
**ANTHROPIC PBC**
(Additional Counsel on Next Page)

**UNITED STATES DISTRICT COURT**

**NORTHERN DISTRICT OF CALIFORNIA**

**SAN JOSE DIVISION**

CONCORD MUSIC GROUP, INC., ET AL.,

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

Case No. 5:24-cv-03811-EKL-SVK

**DEFENDANT ANTHROPIC PBC'S AMENDED AND SUPPLEMENTAL RESPONSE TO PLAINTIFFS' INTERROGATORY NO. 6**

**HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY PORTIONS HIGHLIGHTED IN GREY**

LOUIS W. TOMPROS (*Pro Hac Vice*)
louis.tompros@wilmerhale.com
STEPHANIE LIN (*Pro Hac Vice*)
stephanie.lin@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
60 State Street
Boston, MA 02109
Telephone: (617) 526-6220

ARI HOLTZBLATT (SBN 354361)
ari.holtzblatt@wilmerhale.com
ROBIN C. BURRELL (*Pro Hac Vice*)
robin.burrell@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
2100 Pennsylvania Ave, NW
Washington, DC 20006
Telephone: (202) 663-6000

TAYLOR GOOCH (SBN 294282)
taylor.gooch@wilmerhale.com
KYLE EDWARDS HAUGH (SBN 323952)
kyle.haugh@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
50 California St.
San Francisco, CA 94111
Telephone: (628) 235-1000

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
**LATHAM & WATKINS LLP**
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: (415) 391-0600

SARANG V. DAMLE (*Pro Hac Vice*)
sy.damle@lw.com
**LATHAM & WATKINS LLP**
555 Eleventh Street NW, Suite 1000
Washington, DC 20004
Telephone: (202) 637-2200

ALLISON L. STILLMAN (*Pro Hac Vice*)
alli.stillman@lw.com
**LATHAM & WATKINS LLP**
1271 Avenue of the Americas
New York, New York 10020
Telephone: (212) 906-1747

Pursuant to Rules 26 and 33 of the Federal Rules of Civil Procedure 33 and the Local Rules of this Court, Defendant Anthropic PBC ("Anthropic") hereby serves its amended supplemental objections and response to Plaintiffs' Interrogatory No. 6, served on March 22, 2024. Certain portions of Anthropic's responses are designated "Highly Confidential – Attorneys' Eyes Only" (highlighted in grey), pursuant to the Protective Order entered by this Court. *See* Dkt. 293.

**RESPONSES TO INTERROGATORIES**

**INTERROGATORY NO. 6:**

Identify and describe each dataset used to train Claude, including the dataset's origin; vendor or licensor, if any; number of tokens; contents, including whether the dataset contains song lyrics; and its use in the training process.

**APRIL 22, 2024 RESPONSE TO INTERROGATORY NO. 6:**

Anthropic incorporates its General Objections, Objections to Definitions, and Objections to Instructions as if fully asserted herein. Anthropic objects to this interrogatory as overbroad, unduly burdensome, and disproportionate to the needs of this case because it asks Anthropic to "identify and describe" each and every dataset used to train Claude and is not limited to those datasets that may contain the Works at issue in this case. Anthropic also objects to this Interrogatory because it is compound and consists of multiple subparts, each of which constitutes a separate interrogatory that will count toward the limit of interrogatories that Plaintiffs are permitted to propound. Anthropic further objects to this Interrogatory to the extent it imposes on Anthropic an obligation to provide information that does not currently exist, is unavailable, is only available to be compiled with the expenditure of extraordinary efforts, and/or may be impractical or impossible to create and/or reasonably capture, resulting in a significant, punitive expense to Anthropic. Anthropic also objects to this Interrogatory to the extent that it is cumulative and/or duplicative of other discovery requests by calling for materials or information that Plaintiffs have sought elsewhere, for example in Request for Production No. 28.

Subject to the foregoing General and Specific Objections, Anthropic is willing to meet and

confer regarding the relevance and scope of information sought by this Interrogatory.

**AUGUST 30, 2024 SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 6:**

Subject to the foregoing General and Specific Objections, Anthropic further responds as follows: Anthropic trained the Claude models using publicly available information via the Internet, content obtained from third parties pursuant to commercial agreements, data that its users or crowd workers provided, and data that it generated internally. More specifically, with respect to publicly available data, Anthropic used data that it obtained by crawling public web pages, as well as data it obtained from publicly available Internet sources including Common Crawl. With respect to any data obtained from third parties pursuant to commercial agreements, pursuant to Rule 33(d), Anthropic will produce the relevant agreements, subject to its responses and objections regarding RFP 19. After the agreements are produced, Anthropic will supplement this response to provide the relevant Bates numbers.

Anthropic did not seek out song lyrics to train on. If any song lyrics were included in its training data, it is likely because they are commonly found in numerous places across the Internet, including in an open repository of Internet data archived by Common Crawl.

**DECEMBER 20, 2024 AMENDED SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 6 (HIGHLY CONFIDENTIAL — ATTORNEYS' EYES ONLY):**

Subject to the foregoing General and Specific Objections, Anthropic responds as follows:

Anthropic trained the Claude models using publicly available information on the Internet, nonpublic data from third parties, data provided by data labeling services and paid contractors, and data that it generated internally. More specifically, with respect to publicly available data, Anthropic used data that it obtained by crawling public web pages, as well as data it obtained from publicly available Internet sources including Common Crawl. With respect to any data obtained from third parties pursuant to commercial agreements, pursuant to Federal Rule of Civil Procedure 33(d), Anthropic will produce the relevant agreements, subject to its responses and objections regarding RFP 19. After the agreements are produced, Anthropic will supplement this response to provide the relevant Bates numbers.

Anthropic did not seek out song lyrics to train on. If any song lyrics were included in its training data, it is likely because they are commonly found in numerous places across the Internet. Of 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. However, even for these data sources, it is likely that the overwhelming majority of the data is material other than lyrics:

- Common Crawl and other web crawl collections, such as OpenWebText2
- YouTube subtitles
- Reddit data

Anthropic trains its models on massive amounts of data, including portions of datasets sourced from the Internet (such as, without limitation, components of Common Crawl) and Anthropic does not know all the contents of those datasets. It has made a good faith effort to identify the relevant datasets that are most likely to contain song lyrics. Other datasets may contain song lyrics, but it would be extremely difficult (if not impossible) and enormously burdensome to attempt to forensically analyze those datasets to verify all of the information in them that might include song lyrics or portions of song lyrics.

Anthropic is willing to meet and confer to discuss whether specific information about other datasets that are unlikely to include song lyrics is reasonable and proportional to the needs of the case, including what could be provided, if anything, outside of a training data inspection subject to a Training Data Protocol and/Source Code Protocol to be negotiated in good faith by the parties.

Anthropic's investigation and discovery is ongoing, and Anthropic will supplement its response to this Interrogatory as appropriate.

**SEPTEMBER 11, 2025 AMENDED SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 6 (HIGHLY CONFIDENTIAL — ATTORNEYS' EYES ONLY):**

Subject to the foregoing General and Specific Objections, Anthropic incorporates herein its prior responses to Interrogatory No. 6. In addition, Anthropic responds as follows:

Anthropic further responds that, pursuant to Federal Rule 33(d), it has agreed to make

available non-privileged documents for Plaintiffs' review within the inspection environment including training data for each general purpose generative AI LLM meeting the definition of Claude for which pretraining began on or before November 26, 2024, as described in Response to Plaintiff's Interrogatory No. 14, pursuant to the Stipulated Order Regarding Training Data and Source Code Inspection Protocol, Dkt. No. 386.

Other datasets, such as LibGen and PiLiMi, were used in other ways, such as the use of LibGen in connection with Anthropic's Content Compare System, as indicated by documents previously produced to plaintiffs. *See e.g.*, Anthropic_0000347810.

**OCTOBER 9, 2025 AMENDED SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 6 (HIGHLY CONFIDENTIAL — ATTORNEYS' EYES ONLY):**

Subject to the foregoing General and Specific Objections, Anthropic incorporates herein its prior responses to Interrogatory No. 6. In addition, Anthropic responds as follows:

Neither the LibGen nor PiLiMi datasets, nor any portions of those datasets, were in the training corpus of any of Anthropic's commercially released LLMs. Although Anthropic does not understand Interrogatory No. 6 to request information relating to datasets used to train non-commercial research LLMs, in an effort to avoid further disputes about how Anthropic may have used LibGen or PiLiMi data, Anthropic states that some data from LibGen and PiLiMi were used in the training corpus of one or more internal, non-commercial models used for research purposes.

Date: October 9, 2025

/s/ *Sonal N. Mehta*

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
**LATHAM & WATKINS LLP**
505 Montgomery Street, Suite 2000
San Francisco, California 94111
Telephone: (415) 391-0600

SARANG V. DAMLE (*Pro Hac Vice*)
sy.damle@lw.com
**LATHAM & WATKINS LLP**
555 Eleventh Street NW, Suite 1000
Washington, DC 20004
Telephone: (202) 637-2200

ALLISON L. STILLMAN (*Pro Hac Vice*)
alli.stillman@lw.com
**LATHAM & WATKINS LLP**
1271 Avenue of the Americas
New York, New York 10020
Telephone: (212) 906-1747

SONAL N. MEHTA (SBN 222086)
sonal.mehta@wilmerhale.com
ALLISON BINGXUE QUE (SBN 324044)
allison.que@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
2600 El Camino Real, Suite 400
Palo Alto, CA 94306
Telephone: (650) 858-6000

LOUIS W. TOMPROS (*Pro Hac Vice*)
louis.tompros@wilmerhale.com
STEPHANIE LIN (*Pro Hac Vice*)
stephanie.lin@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
60 State Street
Boston, MA 02109
Telephone: (617) 526-6220

ARI HOLTZBLATT (SBN 354361)
ari.holtzblatt@wilmerhale.com
ROBIN C. BURRELL (*Pro Hac Vice*)
robin.burrell@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
2100 Pennsylvania Ave, NW
Washington, DC 20006
Telephone: (202) 663-6000

TAYLOR GOOCH (SBN 294282)
taylor.gooch@wilmerhale.com
KYLE EDWARDS HAUGH (SBN 323952)
kyle.haugh@wilmerhale.com
**WILMER CUTLER PICKERING HALE AND DORR LLP**
50 California St.
San Francisco, CA 94111
Telephone: (628) 235-1000

*Attorneys for Defendant*
**ANTHROPIC PBC**

**CERTIFICATE OF SERVICE**

I hereby certify that on October 9, 2025 I did cause Defendant Anthropic PBC's Amended and Supplemental Response to Plaintiffs' Interrogatory No. 6 to be served on the following listed below and in the manner so indicated.

**By Electronic Mail**

OPPENHEIM + ZEBRAK, LLP
Matthew J. Oppenheim
Nicholas C. Hailey
Audrey Adu-Appiah (admitted *Pro Hac Vice*)
Corey Miller
4530 Wisconsin Avenue, NW, 5th Floor
Washington, DC 20016
Telephone: (202) 480-2999
matt@oandzlaw.com
nick@oandzlaw.com
aadu-appiah@oandzlaw.com
corey@oandzlaw.com

COBLENTZ PATCH DUFFY & BASS LLP
Jeffrey G. Knowles
Christopher J. Weiner
Bina Patel
Thomas A. Harvey
One Montgomery Street, Suite 3000
San Francisco, CA 94104
Telephone: (415) 772-5795
Ef-jgk@cpdb.com
cwiener@coblentzlaw.com
bpatel@coblentzlaw.com
Ef-Tah@cpdb.com

Jennifer Pariser
Andrew Guerra
Timothy Chung (admitted *Pro Hac Vice*)
Michelle Gomez-Reichman
Bret Matera (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
MGomez-Reichman@oandzlaw.com
Bmatera@oandzlaw.Com

COWN, LEIBOWITZ & 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

Dated: October 9, 2025

By: */s/ Sonal N. Mehta*
Sonal N. Mehta

**VERIFICATION**

I have read the foregoing Anthropic PBC's Amended Supplemental Responses to Plaintiffs' Interrogatory No. 6 and certify under penalty of perjury on behalf of Anthropic PBC that the factual representations in Anthropic PBC's response is, subject to inadvertent and undiscovered errors, true and correct to the best of my knowledge based upon the information made available to me and/or derived from company personnel with knowledge of the matters addressed.



Executed on: October 9, 2025

Jared Kaplan
Co-Founder and Chief Science Officer
Anthropic PBC