EXHIBIT 1

(FILED UNDER SEAL)

1

IN THE UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL., Plaintiffs, -vs- ANTHROPIC PBC, Defendant.) Case No. 5:24-cv-03811-EKL SVK

HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY

VIDEOTAPED DEPOSITION OF:

SHAWN SHAN, Ph.D.

The videotaped deposition of SHAWN SHAN, Ph.D., called as a witness pursuant to notice and pursuant to the provisions of the Federal Rules of Civil Procedure 30 and the Rules of the United States Supreme Court thereof pertaining to the taking of depositions; taken before ISAIAH ROBERTS, CSR, RPR, a Certified Shorthand Reporter, at Arnold & Porter, 70 West Madison Street, Suite 4200, Chicago, Illinois 60602-4231 on March 3rd, 2026, commencing at 9:00 a.m. Central Time.

2

APPEARANCES:

OPPENHEIM + ZEBRAK, LLP
By: MR. KEITH HOWELL
MR. TIMOTHY CHUNG (Appearing remotely)
4530 Wisconsin Avenue
5th Floor
Washington, D.C. 20016
Phone: 202.957.5043
Email: Khowell@oandzlaw.com

Appearing on behalf of the Plaintiffs;

WILMERHALE
By: MS. SONAL MEHTA
MS. ELLEN STEPHANO (Appearing remotely)
2600 El Camino Real
Suite 400
Palo Alto, California 94306
Phone: 628.235.1026
Email: Sonal.mehta@wilmerhale.com

Appearing on behalf of the Defendant.

ALSO PRESENT:

MR. IAN CHEN, Anthropic, In-house Counsel

3

INDEX

PAGE
WITNESS: SHAWN SHAN, Ph.D.
Examination by Ms. Mehta. 5

EXHIBITS

PAGE
Shan Exhibit Number 1. 18
(Expert Report of Dr. Shawn Shan)
Shan Exhibit Number 2. 203
(Rebuttal Expert Report of Dr. Shawn Shan)
Shan Exhibit Number 3. 265
(Claude_inconsistency_test.csv)
Shan Exhibit Number 4. 270
(Claude_transcriptions_coldplay_style.csv)
Shan Exhibit Number 5. 275
(Special_character_prompts_responses_bypassed_run1.csv)
Shan Exhibit Number 6. 280
(Claude_extraction_test_full.csv)
Shan Exhibit Number 7. 298
(Reply Report of Dr. Shawn Stan)

4

THE VIDEOGRAPHER: Here begins Media Number 1 in the videotape deposition of Dr. Shawn Shan in the matter of Concord Music Group, Inc., et al., v. Anthropic PBC in the United States District Court, Northern District of California, San Jose Division, case number 24-cv-03811-EKL-SVK.

Today's date is March 3rd, 2026. The time on the video monitor is 9 a.m. Central. The videographer today is Richard Figueroa representing Planet Depos. This video deposition is taking place at Arnold & Porter, Chicago, Illinois.

Would counsel please voice identify themselves and state whom they represent.

MS. MEHTA: Good morning. Sonah Mehta from WilmerHale on behalf of Anthropic. With me on the Zoom is my colleague Ellen Stephano. And Ian Chen, in-house counsel at Anthropic, is also on the line, on the Zoom.

MR. HOWELL: Keith Howell from Oppenheim + Zebrak on behalf of the music publisher plaintiffs.

THE VIDEOGRAPHER: The court reporter today is Isaiah Roberts, representing Planet Depos. The witness will now be sworn in.

MS. MEHTA: Actually sorry. Just one thing.

Is there anyone else on the Zoom that wasn't announced? I think Tim might be on.

5

THE VIDEOGRAPHER: Timothy Chung is on Zoom.
MR. CHUNG: Timothy Chung on behalf of publishers.
MR. HOWELL: There you go.
SHAWN SHAN, Ph.D.,
called as a witness herein, having first been duly sworn, was examined and testified as follows:
DIRECT EXAMINATION
BY MS. MEHTA:
Q. Good morning, Dr. Shan.
A. Good morning.
Q. Have you ever been deposed before?
A. No.
Q. Okay. Let's go through background rules just to make sure you and I are on the same page about how we'll conduct the day today. Okay?
A. Yes.
Q. Okay. The first thing is we have a written transcript of the testimony, so we need to make sure that we are giving verbal responses. Shaking of the head or nodding or uh-huhs or huh-uhs will not get transcribed on the transcript, so I'd ask you to do your best to please make sure you answer verbally. Okay?
A. Okay.
Q. For the same reason, it's important that you

6

and I not speak over one another, and that you not speak over your counsel. So I'm going to do my best to let you finish your answer before I ask a question. I'm going to ask you to do your best to let me finish a question before you answer. Okay?
A. Okay.
Q. Is there any reason that you wouldn't be able to provide your most truthful and accurate testimony here today?
A. No.
Q. You understand you're testifying under oath just as you would in front of a judge and a jury in San Jose?
A. Yes.
Q. If at any point you don't understand my question, please ask me for clarification. If you don't ask for clarification, I'm going to assume you understood the question, you're providing your most truthful and accurate testimony in response. Okay?
A. Okay.
Q. We'll take a break every hour, 90 minutes or so, just whenever we're at a good stopping place. If you need a break in less time than that, let me know and I'll wrap up that line of questioning and then we'll take a break. Okay?

7

A. Okay.
Q. When were you first approached in this matter?
MR. HOWELL: Objection. Vague.
THE WITNESS: I recall September last year or maybe August last year. I got an email from one of the counsels about retaining me for this matter.
BY MS. MEHTA:
Q. And when you say September or August last year, you mean 2025?
A. That's correct.
Q. Okay. And this was counsel from the Oppenheim + Zebrak firm?
A. That's correct.
Q. Do you know how they got your name?
MR. HOWELL: Objection. Form. Speculation.
THE WITNESS: I do not know.
BY MS. MEHTA:
Q. Had you spoken with Dr. Zhao about this case at any point prior to the lawyers from Oppenheim + Zebrak reaching out to you?
A. I do not recall.
Q. You were working with Mr. Zhao last year; correct?
MR. HOWELL: Objection. Form.
THE WITNESS: He is my advisor during my Ph.D.

8

BY MS. MEHTA:
Q. When did you first start working with him?
MR. HOWELL: Objection. Form.
THE WITNESS: Can you clarify what you mean by working with him?
BY MS. MEHTA:
Q. When did you first -- when did he first become your advisor?
A. When I started my Ph.D., which is 2020, September.
Q. Okay. And have you also worked with him on projects outside of his advisory capacity as you did for your Ph.D.?
A. No.
MR. HOWELL: Objection. Form.
BY MS. MEHTA:
Q. Did you work with him on the Glaze project?
A. Yes.
Q. Okay. So when I had just asked you whether you worked with him on projects outside of your -- his advising on your Ph.D., would you be Glaze to be part of his advising on your Ph.D.?
A. Yes.
Q. Okay. When you started working with him, were there any other apps or tools or projects that you

225

Q. Great. Is it your testimony that you can definitivelyclaim that in generating a particular output Claude references a particular piece of training data?
MR. HOWELL: Objection. Form. Vague. Calls for speculation.
THE WITNESS: Can you repeat that?
BY MS. MEHTA:
Q. Yes. Can you definitively claim that in generating a particular output Claude references a particular piece of training data?
MR. HOWELL: Objection. Form. Vague. Calls for speculation.
THE WITNESS: As we discussed, I do not know the exact underlying of what weights correspond to what memorized training data.
But based on the empirical results and the research that I cited in paragraph 13, it seems to suggest that given a sequence of word -- Claude models oftentime or sometimes will predict mirror definitely next word in a sequence. And from my tests that -- I showed that we can do this for copyrighted lyric, and the probability and the results of this near verbatim extraction is beyond what can happen by random chance.

226

So I believe the training data during the training process which are used to optimize this model have a significant impact on the model's ability to regurgitate these training analyses.
BY MS. MEHTA:
Q. Okay. My question is different than that.
I'm not asking whether you think that it is -- it's suggested or you can infer it.
I am asking whether you can claim definitively that when Claude outputs a particular output, that it is referencing a particular piece of training data?
MR. HOWELL: Objection. Form. Vague. Calls for speculation.
THE WITNESS: What do you mean by referencing specific training data?
BY MS. MEHTA:
Q. That it is retrieving a specific piece of training data. Can you claim that definitively?
MR. HOWELL: Objection. Form. Vague. Calls for speculation.
THE WITNESS: So given how LLMs are trained or how general AI models are trained, the training process, the model weights are changing. They're influenced by the training data pieces. So each individual training piece have some impact to the model weights.

227

And the training models -- as during generation time, the model weights directly impacts what information was getting generated, getting responded. And because model weights are influenced by training data, it is standard to believe training data has a large impact to the final model predictions.
And there are a lot of research looking at -- for example, including some of my research, looking at, for example, given the model generation. Try to explain why model generate this way. And these research are able to show the specific training data points to explain why a model generate this way.
And oftentime it is because similar training data, it's because, you know, certain anomaly in a training set. So I believe a model generation is impacted by this training data.
BY MS. MEHTA:
Q. You did not analyze the model weights used in the Claude models at issue in this case; correct?
MR. HOWELL: Objection. Form.
THE WITNESS: I was not produced with the models weights found in these models.
BY MS. MEHTA:
Q. Okay. Now, going back to your answer about how the training data can influence -- how the model

228

weights are influenced by the training data and therefore the training data has an impact on the final model predictions. That's -- that's about the training data in general. I'm asking about particular pieces of training data.
So can you say definitively that when the model outputs a particular output, that it is retrieving or referencing a particular piece of training data as opposed to the training data as a whole that ultimately is weighted -- that goes into the training mix that ultimately is weighted that ultimately impacts the generation of the output -- or the prediction of the output?
MR. HOWELL: Objection. Form. Vague. Calls for speculation.
THE WITNESS: So there are a lot of research, as I discussed before, that can what they call trace back or understand the influence function of what is a subset of training data points that impact the model to make certain predictions.
And this is useful to, you know, analyze certain behaviors or model to understand what are the causes. And there are very strong analytical background, analytical support for these type of analyses.

229

And to your question, for LLM generation, oftentime it is not one specific data point impacting entire generation. It is, you know, a portion of all related or nearby-related, you know, words and data to it.

And in the case -- and obviously different training data have different impact to certain generation. For example, if you train on Japanese, it's not going to help too much on generally English and vice versa.

And in this case what I was looking at is generating near-verbatim copy of copyright lyrics. And the only place that the copyright lyrics appeared is in the copyright lyrics and the training data. And based on my inspection, those are, you know, hundreds of thousands of copies in there.

So, yes, I believe these -- the presence of these data cause the model to have certain weights. While we cannot understand exactly how these weights are learned, but it nevertheless shows it's able to generate the near-verbatim copy, which is direct consequence of data being trained into the model.

BY MS. MEHTA:

Q. And then when you say we cannot understand how the models are learned, that is -- you mean the -- how

230

the model weights are used to train the model?

MR. HOWELL: Objection. Form. Vague.

THE WITNESS: I'm simply referring to if you show the model weights to someone, it's not human readable; but, again, nevertheless, you know, storing data doesn't need to be in human readable form.

BY MS. MEHTA:

Q. Right. My question was -- oh, actually, never mind. You answered my question. Okay. Thank you. I appreciate that.

Let's go back to paragraph number 10. Let's go -- actually, let's go to paragraph number 20.

A. On page 5?

Q. Yes. And you can just read paragraph 20 to yourself.

A. Okay. I have finished reading paragraph 20.

Q. Okay. Paragraph 20 you talk -- or you respond to Dr. Bilmes' opinion about URLs that contain plaintiff's lyrics, and then you mention or you describe the difference between the absence of a technical block with affirmative legal position -- permission.

Do you see that?

MR. HOWELL: Objection. Form.

MS. MEHTA: Second sentence.

THE WITNESS: Yes, I see that.

231

BY MS. MEHTA:

Q. You're not offering any opinions about what does and does not constitute legal permission to scrape a website, are you?

MR. HOWELL: Objection. Form. Vague.

THE WITNESS: I'm not --

MR. HOWELL: Calls for legal conclusion.

THE WITNESS: I'm not offering legal opinions about what constitute as consenting in a legal sense or not.

But as a general person who owns certain intellectual property rights and have talked -- someone who have spoke to many copyright holders, that is my position, that Robots.txt does not -- is not the same or is not the legal copyright mechanism that right's owners use to manage their copyrights, including myself, on the internet.

BY MS. MEHTA:

Q. Do you know what -- actually, strike that. Let me ask the question this way.

Did you see any evidence in the record in this case about the particular lyric's websites that you reference as being a source for training data and whether or not they engaged Robots.txt?

MR. HOWELL: Objection. Form. Vague.

232

Compound.

MS. MEHTA: I can -- I can ask a better question. That wasn't a very clear question.

BY MS. MEHTA:

Q. There are a number of different lyric websites which you contend are a source for the works in suit lyrics in Anthropic's training data.

Are you with me so far?

A. Yes.

MR. HOWELL: Objection. Form.

BY MS. MEHTA:

Q. Did you see any evidence in this case about whether any of those lyrics websites -- the people that run those websites -- either did or did not intend to authorize Anthropic to -- or anyone else to scrape their websites? Not your speculation, but any specific evidence in this case.

MR. HOWELL: Objection. Form. Vague. Calls for speculation.

THE WITNESS: So based on my analysis of how Anthropic collected data, the data collection happens across many years, including sourcing from Common Crawl and the, you know, data sharks, which was Common

[redacted]

349

songs. I believe the analysis, based on what I reviewed in the research paper were -- which were from respected institutes, respected researchers, they did comprehensive analysis of the internet data.

And based on my internal review of the likelihood of Claude songs appears in the internet data -- and, again, I believe it would be an understatement to say that some what I've tested is likely that many people across these analysis has tested multiple times during their analysis of the internet data of similar type of prompts for verbatim extraction.

BY MS. MEHTA:

Q. For the 499 works in suit?

MR. HOWELL: Thank you, Dr. Shan, for your time.

MS. MEHTA: You're not going to let him answer that question just to -- I'm clarifying the answer he gave.

MR. HOWELL: I think we're three minutes over time now. I've given you a lot of follow-ups.

The deposition's over at this point.

Thank you, Dr. Shan.

MS. MEHTA: Okay. Are you -- are you going to have any questions?

MR. HOWELL: No, I don't have any questions.

350

Thank you for your time, Dr. Shan.

MS. MEHTA: Before we go off the record, I want to mark the transcript highly confidential, attorneys' eyes only.

MR. HOWELL: That's a good point.

And I -- obviously we'll reserve the right to have Dr. Shan review the transcript and make any edits.

THE STENOGRAPHER: I'm sorry. What was that?

MR. HOWELL: Oh, I reserve the right for Dr. Shan to review the transcript and make any edits that -- to correct any statements of his testimony.

THE VIDEOGRAPHER: Stand by. This marks the end of the deposition. We're going off the record. The time is 5:30 p.m.

(5:30 p.m., deposition concluded.)

351

IN THE UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL.,
Plaintiffs,
-vs-
ANTHROPIC PBC,
Defendant.

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

I, SHAWN SHAN, Ph.D., state that I have read the foregoing transcript of the testimony given by me at my deposition on March 3rd, 2026, and that said transcript constitutes a true and correct record of the testimony given by me at said deposition except as I have so indicated on the errata sheets provided herein.

_____
SHAWN SHAN, Ph.D.

No corrections (Please initial)_____
Number of errata sheets submitted_____(pgs.)
SUBSCRIBED AND SWORN to before me this____ day of _____, 2026.

_____
NOTARY PUBLIC

352

CERTIFICATE OF REPORTER

I, ISAIAH ROBERTS, a Certified Shorthand Reporter (IL) and Registered Professional Reporter, being empowered to administer oaths and affirmations remotely pursuant to Section 30(f)(1) of the Federal Rules of Civil Procedure, do hereby certify:

That the foregoing proceedings were taken remotely before me at the time and place herein set forth; that any witness in the foregoing proceedings, prior to testifying, were placed under oath; that a verbatim record of the proceedings was made by me using machine shorthand which was thereafter transcribed under my direction; further, that the foregoing is an accurate transcription thereof.

I further certify that I am neither financially interested in the action nor a relative or employee of any attorney or any of the parties.

Further, that if the foregoing pertains to the original transcript of a deposition in a Federal Case, before completion of the proceedings, review of the transcript was requested.

IN WITNESS WHEREOF, I have this date subscribed my name.

Isaiah Roberts, CSR, RPR
Illinois CSR #084-00489

No. 621066

Re: Deposition of **Shawn Shan, Ph.D.**
Date: 3/3/2026
Case: Concord Music Group -v- Anthropic
Return to: transcripts@planetdepos.com

ACKNOWLEDGMENT OF DEPONENT

I, Shawn Shan, Ph.D., do hereby acknowledge that I have read and examined the foregoing testimony, and the same is a true, correct and complete transcription of the testimony given by me and any corrections appear on the attached Errata sheet signed by me.

4/10/2026 _____

(Date) (Signature)

No. 621066

Re: Deposition of Shawn Shan
Date: 3/3/2026
Case: Concord Music Group -v- Anthropic
Return to: transcripts@planetdepos.com

Page	Line	Correction/Change & Reason
40	14	From "venerated" to "generated"; typo
43	12	From "verify the accuracy" to "verify with accuracy"; transcription error
45	15	From "something is accurate compared to" to "something is AI generated accurately compared to"; transcription error
72	22	From "date" to "data"; typo
82	24	From "ecstatic" to "aesthetic"; transcription error
95	3	From "source guard" to "source code"; transcription error
95	8	From "be done do improve" to "be done to improve"; transcription error
99	5	From "house" to "mouse"; transcription error
107	14	From "have integrated into perfect model" to "have integrated that into a model"; transcription error
109	10	From "inspected those date" to "inspected those data"; typo
118	14	From "me of the inspections" to "me on the inspections"; typo
147	6	From "incidence" to "incidents"; transcription error
161	2	From "doe" to "code"; typo
163	24	From "even have the exact date" to "even having the exact data"; transcription error
187	22	From "a common message" to "a commit message"; transcription error
189	13	From "Git common message" to "Git commit message"; transcription error
189	14	From "put Git common" to "put Git commit"; transcription error
215	17	From "ever" to "every"; typo
223	3	[redacted] typo
244	4	From "correctly checks" to "correctly checked"; transcription error
245	10	From "color" to "code"; transcription error
255	14	From "glasses" to "classes"; typo
262	13	From "hear" to "here"; typo
291	14	From "Cleo" to "Clio"; transcription error
291	18	From "no" to "not"; typo
297	2	From "Cleo" to "Clio"; transcription error
299	10	From "produced to be" to "produced to me"; typo
302	4	From "denotes" to "denoted"; typo
312	12	From "it's rarely have false" to "it rarely has"; transcription error
312	14	From "copyrighted" to "copyright"; typo
313	22	From "developer typically do" to "developer will typically do"; transcription error
317	4	From "terminal" to "terms of"; transcription error
319	16	From "CM" to CMI"; typo

321	15	From "include removes removing" to "include removing"; transcription error
327	5	From "that are the cases I seen" to "those are the cases I've seen"; transcription error
345	6	From "nd the fact" to "And the fact"; typo

4/10/2026
(Date)

(Signature)