EXHIBIT 4

* UNCERTIFIED ROUGH DRAFT *

This transcript is an UNCERTIFIED ROUGH DRAFT TRANSCRIPT. It contains raw output from the Court Reporter's stenotype machine translated into English by the Court Reporter's computer without the benefit of proofreading. It will contain untranslated steno outlines, mistranslations (Wrong Words), and misspellings. These and any other errors will be corrected in the final transcript.

Since this ROUGH DRAFT TRANSCRIPT has not been proofread, the Court reporter cannot assume responsibility for any errors therein. This ROUGH DRAFT TRANSCRIPT is intended to assist attorneys in their case preparation and is not to be construed as the final transcript. It is not to be read by the witness or quoted in any pleading or for any other purpose and may not be filed with any court.

* * *

THE VIDEOGRAPHER: Here begins media number 1 in the videotaped deposition of Benjamin Mann, individual, in the matter of Concord Music Group v. Anthropic, in the United States District Court, Northern District of California, San Jose Division, case number 24-CV-03811-EKL. Today's date is November 20, 2025. The time on the video monitor is 16:06.

The videographer today is Phillip Astor, representing Planet Depos. This video deposition is taking place at Wilmer Cutler Pickering Hale and Dorr LLP, San Francisco, California.

Will counsel please voice-identify themselves and state whom they represent?

MR. MILLER: Corey Miller from Oppenheim & Zebrak, LLP on behalf of the plaintiff music -- plaintiff music publishers. And also here with me from Oppenheim & Zebrak is Michelle Gomez-Reichmann.

MS. MEHTA: Good afternoon. Sonal Mehta from WilmerHale on behalf of Anthropic and the witness. With me today is my colleague Taylor Gooch. We also have on the Zoom our colleague Jared Grubow. We have with us in-house counsel from Anthropic in the room, Devon Hanley Cook, and

we also have Sara Sampoli, our co-counsel from Latham & Watkins.

THE VIDEOGRAPHER: The court reporter today is April Crites, representing Planet Depos.

The witness will now be sworn.

BENJAMIN MANN

of lawful age, Defendant herein, having been first duly cautioned and sworn, as hereinafter certified, was examined and said as follows:

CROSS-EXAMINATION

BY MR. MILLER:

Q Good afternoon, Mr. Mann.

Do you understand that you are under oath?

A Yes, I do.

Q And that's the same oath to tell the truth you would take if we were in a courtroom before a judge and jury?

A Yes, I understand that.

Q Is there any reason you cannot testify today in accordance with that oath?

A There's no such reason.

Q Okay. Have you been deposed before, Mr. Mann?

A Yes, I have.

Q How many times?

THE VIDEOGRAPHER: Back on the record at 16:40.

MR. MILLER: Tab 29.

(Exhibit 202, Description, was marked and presented for purposes of identification.)

BY MR. MILLER:

Q Mr. Mann, in the Zoom, you should see what's been marked as Plaintiffs' Exhibit 202.

Does 202 appear to be a copy of a paper titled, Language Models are Few-Shot Learners?

THE COURT REPORTER: Can you repeat that one title for me, please?

MR. MILLER: Sure. Language Models are Few-Shot Learners.

THE WITNESS: Just to confirm, this is not intended to have a Bates number?

BY MR. MILLER:

Q Correct.

A Okay. Yes, it does appear to be a copy of that paper.

Q And you're listed as one of the co-authors of this paper, correct?

A Yes, I'm one of the four first authors.

Q And this is a research paper you worked on while you were employed at OpenAI?

A That's correct.

Q What was your contribution to this paper?

A Well, there's the contribution section close to the end of the paper that details exactly what my contributions were. But my primary contributions were around acquisition, processing, filtering, and experiments to validate the data mixes for the models, as well as a few architecture experiments.

Q Did you decide what data -- withdrawn.

When you said your contributions were around acquisition, processing, filtering, and experiments to validate the data mixes for the models, what are the models you're referring to?

A There were many different research models that we made.

Q Mm-hmm.

A But eventually, we created a model that externally had the name GPT-3.

Q And did you decide what data to collect for use in training GPT-3?

MS. MEHTA: Objection to the form.

Mr. Mann, again, I'm going to caution you about any confidentiality obligations you have with respect to OpenAI. So anything confidential, you

cannot share. But if there is public information about that's responsive to the question, you can share that public information only.

THE WITNESS: I don't think there's public information about who made the decisions on what data would be included in the training mix.

BY MR. MILLER:

Q Can you turn to page 9 of the exhibit, please?

A Okay.

Q And you see in the middle of page 9, there's a -- there's a chart?

A I see a chart titled Total Compute Use During Training.

Q So I'd direct your attention to the table beneath that, which I think is labeled table 2.2, that's described as datasets used to train GPT-3.

Do you see that?

A Yes, I do.

Q And does the table directly above that list the datasets used to train GPT-3?

MS. MEHTA: Objection to the form.

And, again, Mr. Mann, if you can answer -- if the question can be answered based on public information, you can answer it. If the question

is calling for confidential OpenAI information, then I caution you not to answer that in light of your confidentiality obligations.

THE WITNESS: So the GPT-3 model that was described in the paper, yes, these are the datasets that were used to train that model.

BY MR. MILLER:

Q And the first dataset listed in table 2.2 is Common Crawl, correct?

A Yes, that's a shorthand for a dataset that we derive from the Common Crawl data.

Q And the second dataset listed in table 2.2 is WebText 2.

Do you see that?

A Yes.

Q What is WebText 2?

A WebText 2 is, I believe, described elsewhere in the paper, and it is the second version, an expanded version of what was described, I believe, in the GPT-2 paper as WebText 1, which, at the time, since there was no 2, it was just called WebText, which is a, at the time, large-scale corpus of natural language text sourced from web pages that were back-linked from Reddit posts with at least some minimum score of

karma.

Q The next dataset on the list is Books1.

Do you see that?

A Yes, I do.

Q What is Books1?

MS. MEHTA: Form.

THE WITNESS: Books1 was a dataset that we used to train GPT-3.

BY MR. MILLER:

Q And what did Books1 consist of?

MS. MEHTA: Same objection.

THE WITNESS: It's not public information what was contained in Books1.

BY MR. MILLER:

Q Is Books1 a dataset that OpenAI created?

MS. MEHTA: Objection to the form.

Again, Mr. Mann, anything that is public you can share. If it is confidential to OpenAI, you cannot share it.

Also, Corey, I have no idea what possible relevance all these questions about what OpenAI was and wasn't doing could have to this case, which is about Anthropic. I've given you a lot of leeway, but you're just wasting time at this point.

THE WITNESS: Can you restate the

question?

BY MR. MILLER:

Q Is Books1 a dataset that OpenAI created?

MS. MEHTA: Same caution: If it's public, you can share it. If not, if it's confidential to OpenAI, you cannot.

THE WITNESS: I don't think I can say whether Books1 was or was not a dataset that OpenAI created. But I can say that I prepared the training data for the GPT-3 training run.

BY MR. MILLER:

Q Are you familiar with Dario Amodei?

A Yes, I am.

Q And he's Anthropic's CEO, right?

A Yes.

Q How did you first meet Mr. or Dr. Amodei?

A I first met Dario I believe at OpenAI. The first time that I worked there in 2017.

Q Do you work closely with Dr. Amodei at Anthropic?

A At this point, I would say I don't work closely with him. He has a pretty different job than he used to have when I did work closely with him.

Q Was there a time at Anthropic when you

worked closely with him?

A Yes.

Q When was that?

A I would say when we first started the company and we were building our very first models, he was quite involved in the research process, and then subsequently, when I convinced the company that we should start a product and product engineering team, Dario was pretty closely involved with our product plans, and we had regular meetings to discuss what we should do.

Q And then at some point, did Dario -- excuse me, did Dr. Amodei's degree of involvement change?

A The concrete thing that happened is I left product engineering leadership. So I created the team, and I built it up, but at some point, we hired Brian Delahunty to be our head of project engineering. And at that time, I went off to start a brand-new team that wasn't working closely with Dario anymore.

Q Does Dr. Amodei play any role with respect to acquiring datasets to be used to train Claude models?

MS. MEHTA: Objection to the form.