EXHIBIT 125



Planet Depos®
We Make It Happen™

HIGHLY CONFIDENTIAL
ATTORNEYS' EYES ONLY

Transcript of Benjamin Mann

Date: November 20, 2025
Case: Concord Music Group -v- Anthropic

Planet Depos
Phone: 888.433.3767
Email: transcripts@planetdepos.com
www.planetdepos.com

1

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

CONCORD MUSIC GROUP, INC., et al., : Civil Action No. 5:24-cv-03811-EKL-SVK

Plaintiffs,

v.

ANTHROPIC PBC,

Defendant.

HIGHLY CONFIDENTIAL
ATTORNEYS' EYES ONLY

VIDEOCONFERENCED (HYBRID) AND VIDEO-RECORDED

DEPOSITION OF BENJAMIN MANN

PURSUANT TO FEDERAL RULE OF CIVIL PROCEDURE 30(b)(1)

San Francisco, California

Thursday, November 20, 2025

4:06 p.m. Pacific Standard Time

VOLUME I

Job No. 609867

Pages: 1- 145

Reported By: April L. Crites, RMR, CRR, CSR, CCR

California Certified Shorthand Reporter

2

DEPOSITION OF BENJAMIN MANN

HELD AT THE OFFICES OF:

WILMER CUTLER PICKERING HALE and DORR, LLP

50 California Street

Suite 3600

San Francisco, California 94111

(628) 235-1000

* * *

Pursuant to notice, and pursuant to the Federal Rules of Civil Procedure, before April L. Crites, RMR, CRR, CSR, CCR, and California Certified Shorthand Reporter No. 14493.

* * *

3

I N D E X

EXAMINATIONS CONDUCTED	PAGE
BENJAMIN MANN	
Cross-Examination by Mr. Miller.....	10

* * *

EXHIBITS

MARKED	DESCRIPTION	PAGE
Exhibit 87	Execution: Copyfairy as a sampling stage, Database, Bates stamped Anthropic_0000429077 through Anthropic_0000429094	106
Exhibit 183	[redacted] Launch signoff (Product, Models, T&S, Comms, Bates stamped Anthropic_12437 through Anthropic_0000012449	124
Exhibit 200	Benjamin Mann's LinkedIn web page	12
Exhibit 201	ANTHROP\C org chart, Bates stamped Anthropic_0000000168	23
Exhibit 202	arXiv:2005.14165v4 [cs.CL] 22Jul2020, Language Models are Few-Shot Learners, OpenAI	26
Exhibit 203	January 5, 2021, e-mail from Notion to [redacted] Bates stamped Anthropic_0000458349 through Anthropic_0000458353	47
Exhibit 204A	Excel spreadsheet: The Below are TBD Pending Pretraining Team Discussions Post Scan Launch, Bates stamped Anthropic_0000458418	65
Exhibit 204B	Excel spreadsheet, "Note: Ben has a similar sheet here; dataset name, Bates stamped Anthropic_0000510804	70
Exhibit 205	Slack communication, dated June 4, 2021, Bates stamped Anthropic_0000014096 through Anthropic_0000014106	88
Exhibit 206	January 4, 2024, e-mail from Sam Martin to [redacted] Subject: Dataset for copyright filter checks, with attachment, Bates stamped Anthropic_0000218330	108
Exhibit 207	January 4/5, 2024, e-mail from [redacted] to Sam Martin and Ben Mann, Subject: Re: Dataset for copyright filter checks, Bates stamped Anthropic_0000218341 through Anthropic_0000218342	113

4

5

Exhibit 208	January 8, 2024, e-mail from Evan Frondorf, Subject: Upcoming ability to intervene on prompts, Bates stamped Anthropic_0000218344	114
Exhibit 209	DRI: Miranda Zhang, last substantial edit June 27, 2023, Bates stamped Anthropic_0000479446 through Anthropic_0000479474	136

* * *

6

A P P E A R A N C E S

ON BEHALF OF THE PLAINTIFFS, THE MUSIC PUBLISHERS:

Corey Miller, Esq.
Michelle Gomez-Reichman, Esq.
Nicholas C. Hailey, Esq.
OPPENHEIM + ZEBRAK LLP
4530 Wisconsin Avenue NW
Fifth Floor
Washington, D.C. 20016
202-596-5043
corey@oandzlaw.com
mgomez-reichman@oandzlaw.com
nick@oandzlaw.com

ON BEHALF OF THE DEFENDANT, ANTHROPIC, PBC:

Sonal N. Mehta, Esq.
WILMER CUTLER PICKERING HALE & DORR, LLP
2600 El Camino Real
Suite 400
Palo Alto, California 94306
650-858-6000
sonal.mehta@wilmerhale.com

- AND -

7

A P P E A R A N C E S C O N T I N U E D

ON BEHALF OF ANTHROPIC, PBC (CONTINUED):

Joseph Taylor Gooch, Esq.
WILMER CUTLER PICKERING HALE & DORR, LLP
50 California Street
Suite 3600
San Francisco, California 94111
628-235-1002
taylor.gooch@wilmerhale.com

- AND -

Jared V. Grubow, Esq. (Via Zoom)
WILMER CUTLER PICKERING HALEK& DORR, LLP
7 World Trade Center
250 Greenwich Street
New York, New York 10007
212-295-6474
jared.grubow@wilmerhale.com

- AND -

8

A P P E A R A N C E S C O N T I N U E D

ON BEHALF OF ANTHROPIC, PBC (CONTINUED):

Sara E. Sampoli, Esq.
LATHAM & WATKINS LLP
555 Eleventh Street NW
Suite 1000
Washington, D.C. 20004
202-637-2200
saa.sampoli@lw.com

ALSO PRESENT:

Devon Hanley Cook, Esq.
Anthropic PBC

Philip Astor, Videographer
Planet Depos

* * *

Pages 17-28

17

and for me, this was a big indicator of how we could get from where we were at the time to artificial general intelligence, which is to say, AI models that could do many of the things that humans could do.

And the reason that I joined OpenAI in the first place was because I was concerned that as we got closer to AGI, the chances of that AGI becoming misaligned and leading to existential-risk scenarios for -- for humanity could be likely.

And so when GPT-2 came out, for me, this was an indicator that not only should I not abandon AI safety, but instead, I should double down, and it was still the only problem that I deeply cared about at the time.

I had friends who had gone back to OpenAI, like Tom Brown and Nick Cammarata, who told me that things were different, that Elon Musk had left the company, that people actually knew who their managers were, and that gave me confidence that I could work there sustainably; I could have an impact; and that I would feel a sense of purpose.

And so, ultimately, those are the -- the main factors that led to my returning.

18

Q And why did you subsequently leave OpenAI to join Anthropic?

MS. MEHTA: Same caution with respect to anything that would be confidential to OpenAI, but you can answer.

THE WITNESS: When I was in sort of the last couple of months of my time at Anthropic, there were multiple instances where the safety team, which was the team that I was on, along with the other Anthropic cofounders, and the executive team, which at the time was led by Sam Altman, would disagree about our company strategy, and in particular, we felt that safety wasn't really considered to be a first-class priority.

And I think there are many reasons that you could reasonably believe that, but we all believed that safety was the most important thing, which is not very surprising, given the team that we were on. And the other seven cofounders all decided to leave together. I wasn't involved in those discussions or decisions, and so they -- they sort of all left together.

And before I left, Mira Murati, who, at the time, was, I believe, the CTO of the company,

19

offered for me to become the leader for the safety team, which, at the time, was headless because all the Anthropic cofounders, with the exception of me and Tom Brown, were safety team leads, and the only other one was Paul Christiano, who also decided to leave and start his own company.

And that was actually a very difficult decision for me to make, whether I should stay and try to keep the torch of safety burning at OpenAI, or leave -- (audio interruption) -- and start from zero and work with these people, who I had worked with for the preceding year and a half and trusted very, very deeply, but nonetheless were starting with no funding and no code that we had been developing for quite sometime and no resources, no real reputation, relatively speaking.

And, ultimately, I made the decision to leave because I had a conversation with one of the founders of Open Philanthropy project, who I deeply trusted, and he made an analogy for me that really resonated. And the analogy was, when you're dating someone, either you can go for the person who looks good on paper or you can go for the person who you have chemistry with.

And that just immediately made it became

20

clear that I wanted to go to Anthropic and work with all these people who I really loved working with and deeply, deeply trusted, instead of playing a political game for the next n years of my career.

And on top of that, I also had a conversation with Yann LeCun, who Dario had hired but who hadn't started working at OpenAI yet, and I told him, If you come to this company and you become the lead of the safety team, you're going to be playing a political game. You're not necessarily going to get to focus on the safety research that you're dedicated to doing. Are you okay with that? And it was because he said yes that I felt comfortable sort of following my heart and going to Anthropic.

BY MR. MILLER:

Q Are you -- withdrawn.

Do you consider yourself a cofounder of Anthropic?

MS. MEHTA: Objection to the form.

THE WITNESS: So from a corporate governance standpoint, I am not a cofounder of Anthropic. However, between me and the other cofounders, de facto, I am, and I've played that

21

role since the beginning. From an equity standpoint, I'm not a cofounder, but I do play that role, and so we agreed that I could call myself publicly a cofounder.

BY MR. MILLER:

Q What is your current title at Anthropic?

A My current title is the same as all other people in a somewhat similar capacity, which is member of technical staff.

Q What are your current responsibilities?

A I have many different responsibilities. I call myself the engineer of last resort, which is supposed to mean that if something isn't getting done or there's some conflict between different groups or parts of the organization, then I will jump in and try to resolve it.

But explicitly, I am currently the technical lead for product engineering. I run a very small group of tech leads that sort of does all the same things that I do to extend my reach. The research PM organization reports up to me, which was their request. And I am now rebooting a team that I used to run but that didn't exist for some period of time, which we call labs, which is responsible for doing technology transfer from

22

research to product.

Q You referred to the research PM organization as reporting up to you.

What is the research PM organization?

A "Research PM" refers to research product management. This is a group of people with a product management background that comes from many different places across the industry, and their goal is to ensure that the researchers are hearing the needs and -- and trends in the industry and from customers to prioritize new model capabilities and figure out things like our product hypothesis for new model entries that we might bring to market.

Q So the research PM organization reports up to you.

The labs group that you referred to, do they also report up to you?

A The engineering manager for labs reports to me.

Q Okay.

A But the PM reports to the research PM lead, so indirectly, yes.

Q To whom do you directly report?

A I currently report to Rahul Patil, who is

23

currently our new CTO.

(Plaintiffs' Exhibit 201, ANTHROP\C org chart, Bates stamped Anthropic_0000000168, was marked and presented for purposes of identification.)

BY MR. MILLER:

Q In the Zoom, you should see a document bearing the Bates number Anthropic_618.

MR. MILLER: Do you know if it's the version that's been marked before?

Do you know what number it was previously marked with?

BY MR. MILLER:

Q That, we'll be marking as Plaintiffs' Exhibit 201.

Do you see Anthropic_618 in front of you?

A Yes, I do.

Q And is this exhibit an org chart for Anthropic's leadership?

MS. MEHTA: Objection to the form.

THE WITNESS: It is a version of an org chart, but I would say this is quite out of date at this point.

BY MR. MILLER:

Q Okay. Do you see you're listed on here

24

twice -- once -- I'm sorry -- not twice -- just -- withdrawn.

Do you see, under Products, you're listed as under labs?

A Yes.

Q A minute ago, you said that labs was being rebooted.

Did labs previously exist in some form?

A Yes.

Q What time period, approximately, did labs exist, previously?

A We started labs in April of 2024.

Q And then when did labs end?

A It ended approximately in the first half of 2025.

Q So, at this time, were you reporting to Mike Krieger?

A I would need to see the date to verify, but I did report to Mike Krieger for a significant portion of the times -- of the time that I was running the first version of labs.

Q How does your work at Anthropic differ from what your role was at OpenAI?

MS. MEHTA: I'm going to again caution you not to disclose any confidential information about

25

your work at OpenAI in light of your obligations, but you can answer that to the extent you can do so without revealing that confidential information.

THE WITNESS: At OpenAI, I was an individual contributor for most of my time there, and I was on the data engineering and research side of the house. Towards the end, I was on more a prototyping kind of role, but in general, it was very researchy.

At Anthropic, I have had a very wide variety of jobs, which has come for cofounders, including managing large teams, starting brand-new teams and giving them away, temporarily jumping in for managers for many different reasons, including medical leave, maternity leave, et cetera.

And so I've done things like start and managed the IT and security teams, managed the ops team, started our product and product engineering teams.

And so it would be pretty hard for me to give a comprehensive list of all the things that I have done at Anthropic, just given the variety over the many years that I've worked with the company, but those are some examples.

26

MR. MILLER: All right. Can we take a short break?

MS. MEHTA: Sure.

THE VIDEOGRAPHER: We are going off the record at 16:31.

(A brief recess was taken.)

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

MR. MILLER: Tab 29.

(Plaintiffs' Exhibit 202, arXiv:2005.14165v4 [cs.CL] 22Jul2020, Language Models are Few-Shot Learners, OpenAI, 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:

27

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.

28

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,

Pages 49-52

49

Q Okay. Can you turn to the bottom of the second page, please. I want to direct your attention to the third-to-last paragraph.
Do you see that the document states that, The Pile contains lots of great data sources for us to get started with?
A I see that.
Q Do you know whose words those were?
A They're either my words or Sam McCandlish's words.
Q Would you agree that in 2021, the Pile contained lots of great data sources for what became Anthropic to get started with?
MS. MEHTA: Objection to the form.
THE WITNESS: I think that's a totally reasonable thing to say, yeah.
BY MR. MILLER:
Q Has -- has Anthropic used the Pile or data from -- included in the Pile dataset to train its commercially released AI models?
MS. MEHTA: Form.
THE WITNESS: Yes, we have used subsets of the Pile to train commercially released models.
BY MR. MILLER:
Q Has Anthropic used subsets of the Pile to

50

train all of its commercially released models?
MS. MEHTA: Objection to the form.
THE WITNESS: I don't know about all.
BY MR. MILLER:
Q Does Anthropic still use the Pile to -- or its -- withdrawn.
Does Anthropic still use data from the Pile datasets to train its commercially released models?
MS. MEHTA: Objection to the form.
THE WITNESS: I don't know.
BY MR. MILLER:
Q A little while ago we discussed OpenWebText2.
Do you recall that?
A Yes.
Q Have you personally reviewed any of the subsets of the OpenWebText2 dataset?
A You're referring to the OpenWebText2 dataset that was referenced in the GPT-3 paper?
Q Well, let's take a step back.
There was an OpenWebText -- is there more than one OpenWebText2 dataset?
A Oh, sorry, I had an internal confusion.
So there's WebText2 --

51

Q Mm-hmm.
A -- which was referenced in the GPT-3 paper.
Q Mm-hmm.
A And then there's OpenWebText, but I don't believe there's OpenWebText2, but maybe there is.
Q Okay. It's --
A I don't recall the details.
Q It's possible the confusion is on our -- on my end. So I apologize.
So there's -- are you familiar with a dataset called OpenWebText?
A Yes, I have heard of a dataset that's named something like that.
Q Is there any relationship between the Open WebText dataset and the WebText or WebText2 datasets?
A So when GPT-2 came out and they talked about WebText, there were a number of groups that wanted to reproduce the results in GPT-2. So those groups banded together to produce what was called the OpenWebText. I don't know whether they made an OpenWebText2 or not, but they reproduced the methods described in the GPT-2 and possibly GPT-3 papers to create that dataset.

52

Q Has Anthropic obtained the OpenWebText dataset?
A I don't know, but I'm sure that there are documents furnished for this case that I could use to verify that if you wanted to show them to me.
Q Has Anthropic obtained the WebText2 dataset?
A That would only be possible by agreement with OpenAI, which we certainly don't have.
Q Are you familiar with a YouTube subtitles dataset?
A Yes.
Q What is the YouTube subtitles dataset?
MS. MEHTA: Form.
THE WITNESS: So there are two different YouTube subtitles datasets that I'm familiar with. One is the subset of the Pile that is YouTube site subtitles, and then another one is one that Anthropic acquired. And I believe these datasets contain scraped YouTube subtitles.
BY MR. MILLER:
Q Okay. Is there overlap between the subset of the Pile that contains YouTube subtitles and the separate YouTube subtitles dataset?
MS. MEHTA: Objection to the form.

Pages 81-92

81

things like deduplication --
MR. MILLER: Mm-hmm.
THE WITNESS: -- which is a form of filtering, yes, we don't want so many copies of that data to appear in the final outputs, but that process is agnostic to the value of a single copy or a small number of copies of that data. So I would say there are some filtering techniques that we use that don't bear on whether, quote/unquote, Anthropic wants that data.
Q Are there some filtering techniques Anthropic uses that do bear on whether Anthropic wants that data?
MS. MEHTA: Objection to the form.
THE WITNESS: So you can think of quality filtering as a form of data filtering. And for that, again, we usually [redacted]. So the way it works is we get [redacted]. But back in the day, we used the [redacted] to compute a [redacted]
And then once we had that score between [redacted] we would decide [redacted] whether to keep or not keep that

82

document.
And we would bias towards keeping more of the documents that [redacted] high quality and fewer of the documents [redacted] were low quality, but there was always a small chance that we would keep documents that we thought were low quality and throw out documents we thought were high quality, because we had uncertainty -- or, rather, I would say [redacted]
And if we were to throw out every document that the [redacted] we would likely significantly skew our data distribution in a way that damaged model performance. And we empirically validated that hypothesis, as well.
BY MR. MILLER:
Q Are you familiar with something called an extractor tool?
MS. MEHTA: Objection to the form.
THE WITNESS: There are many tools that we have used over the years that I would classify as extractors.
BY MR. MILLER:
Q What is an extractor?

83

A So, typically, an extractor refers to a piece of software that goes from a more voluminous or uncompressed or otherwise raw form of some data down to a more training-ready form of that data.
So, for example, we would have PDF extractors that took PDF data and turned it into a plain-text representation.
BY MR. MILLER:
Q Has Anthropic created its own extractors, or does it use third-party-created extractor tools?
MS. MEHTA: Objection to the form. Foundation. Speculation.
THE WITNESS: So we have certainly created a lot of extraction software, but it is often built on top of open source technology. But even then, many of these open source libraries contain many different knobs and dials that can be used to tune an extraction process.
And at that point, it's up to us -- empiricism, intuition, spot-checking of the data -- to decide what kind of extraction we should be doing. And sometimes we'll even use multiple different extraction techniques for the same type of content to provide variety.

84

BY MR. MILLER:
Q You mentioned using extractors on PDFs. Has Anthropic also used extractors on HTML documents?
MS. MEHTA: Objection to the form. Foundation.
THE WITNESS: Yes, I think it would be fair to say that.
BY MR. MILLER:
Q For what purpose has Anthropic used extractors on HTML documents?
MS. MEHTA: Objection to the form.
THE WITNESS: Well, a natural HTML document as would you find it on the web contains a large amount of boilerplate -- for example, ad-tracking libraries that are in JavaScript embedded in the page, different kinds of invisible content that's been rendered to a user, site navigation, all kinds of content that won't really help the model learn.
And as we've discussed before, we have pretty strong evidence that duplicated data in model training causes the model to be less capable after training.
And so for all these reasons, it's very

85

valuable for us to get rid of as much of that extraneous content as possible before training the model.
BY MR. MILLER:
Q Does the extraneous content that Anthropic seeks to get rid of also include copyright notices on a web page?
MS. MEHTA: Objection to the form. Calls for speculation.
THE WITNESS: It depends strongly on the web page itself and how that web page is structured. The way most of these content extractors work on HTML is they look for things like

tags inside the main body of the HTML

document, and so if a copyright notice were in the body in a

tag, then I find it highly unlikely

that we would remove it.
BY MR. MILLER:
Q Do you know whether -- withdrawn.
What is a

tag?

A In HTML -- HTML stands for hypertext markup language.
Q Mm-hmm.
A And what that's referring to is there's this structure, language, that contains many

86

different kinds of tags. I mentioned before there's the body tag, there's a

tag

tags

generally are meant to semantically represent paragraphs of text, but they are also used just for delineating one piece of content from another in a web page.
There is a very, very wide distribution of how these kinds of tags are used on the Internet, so it's pretty hard to generalize. But that's sort of what the specification intended

tags

to be used for.
MR. MILLER: Tab 2? Do you remember what number this was marked?
MS. GOMEZ-REICHMAN: 187.
BY MR. MILLER:
Q In the Zoom, you should see what we marked earlier this morning as Plaintiffs' Exhibit 187, which has the Bates number Anthropic_349548.
Do you remember this document?
A Yes, it looks familiar.
Q And I want to you go all the way to the end, very bottom of the last page.
Do you see at the bottom of the page, there's a bunch of links?
A Yes.

87

Q And then, are those links the kinds of things Anthropic would try to remove during the filtering process?
MS. MEHTA: Objection to the form. Calls for speculation.
THE WITNESS: Well, this is generally site navigation, so that is the kind of thing that would typically be removed by our HTML extraction libraries.
BY MR. MILLER:
Q And then underneath the site navigation links, in the bottom left corner, do you see it says Copyright 2025 Anthropic PBC?
A Yes, I see that.
Q Is that sort of copyright notice something Anthropic would typically try to remove using an extractor in the filtering process?
MS. MEHTA: Objection to the form.
THE WITNESS: It's not something that we would specifically try to remove. But these extractors would potentially do it. I -- without trying it on this page, I wouldn't be able to tell you what the effect would be.
BY MR. MILLER:
Q Are you familiar with the Newspaper

88

extractor?
A Yes, I've used it.
Q What is the Newspaper extractor?
A Newspaper is an HTML-to-plain-text converting library. I believe it has a wide variety of other potential uses, but that's how we use it.
Q Okay. Are you familiar with the Readability extractor?
A Yeah.
Q And what is the Readability extractor?
A So there's at least one library called Readability that is maintained by Mozilla, and that also takes web pages and converts them into a simplified HTML or a plain-text representation.
Q Are you familiar with a tool called jusText?
A Yes.
Q And What is jusText?
A It's in the same class of software as Newspaper and Readability.
Q Okay.
MR. MILLER: Okay. Tab 37.
(Plaintiffs' Exhibit 205, Slack communication, dated June 4, 2021, Bates stamped

89

Anthropic_0000014096 through Anthropic_0000014106, was marked and presented for purposes of identification.)

BY MR. MILLER:

Q In the Zoom you should find what's been marked as Exhibit 205, which bears the Bates number Anthropic_14096.

A I see that.

Q Does this appear to be a series of Slack communications with the communication at the top dated June 4, 2021?

A Yes.

Q Can you turn to page 3, please?

And do you see in the middle of the page there's a comment from Ben?

A I see multiple comments from Ben.

Q Ben is you, right?

MS. MEHTA: Objection to the form.

THE WITNESS: Probably me, yes.

BY MR. MILLER:

Q Do you see the -- the comment from you that begins, Done benchmarking extractors?

A Yes.

Q Do you recall doing any work to benchmark extractors?

90

A Yes.

Q What was that work?

A I'm going to read a bit more of the document to make sure that I understand correctly.

Okay. I've reviewed the -- the document.

Can you ask your question again?

Q Yeah. What was the work you did in relation to benchmarking extractors?

MS. MEHTA: Objection to the form.

THE WITNESS: My recollection, based on this thread, is that there were three different things that I was looking for in extraction benchmarking.

One was extraction speed. So that's, given a -- the same document, how fast are the different extractors at producing the output.

The second was extraction error rate. So often the extractors will just produce some error and not produce anything at all, despite having some imported input.

And then the third is quality, and that's a bit harder to measure. But based on the spot-checking, it looks like I and the other people in this thread got a sense of the different quality levels.

91

Q When you refer to extraction error rate, does -- does that refer to the extractors removing too much information and not leaving enough behind, or removing too little information and leaving too much unwanted data behind?

MS. MEHTA: Objection to the form.

THE WITNESS: Actually, neither of those.

BY MR. MILLER:

Q What does -- then what does error rate mean in this context?

MS. MEHTA: Form.

THE WITNESS: So in software engineering, when a piece of code errors, it means that some precondition or assumption in the code was broken, and before finishing execution of the code, the program exits, often with a non-zero exit code, which indicates failure and a message indicating what the failure was.

So, in this case, when I'm talking about error rates, I'm talking about the code basically breaking.

BY MR. MILLER:

Q Could you turn to page 6, please?

A Yep.

Q Do you see your comment in the -- fourth

92

from bottom on this page?

A The one that starts with, It is representative?

Q No. The one that begins, Here's your 2013 comparison.

Do you see that?

A I see that.

Q And then it says, Maybe even more garbage-y.

And there's a link to something.

A Yes.

Q And then underneath it, it says, This is why I'm more excited about LibGen.

Do you see that that?

A Yes.

Q Why did you mean by saying, I'm more excited about LibGen?

MS. MEHTA: Objection to the form.

THE WITNESS: In general, the subject of this thread is finding high-quality data and being able to extract it into a form that we can train our models on.

My concern with this data was that it looked like it didn't have a lot of contiguous high-quality, human-readable text, and my

Pages 101-104

101

There's a tension between building a model that's helpful and building a model that's harmless, right?

A That's correct.

Q An entirely harmless model could potentially be not very helpful for anything, right?

A A rock would be --

MS. MEHTA: Objection to the form.

THE WITNESS: A rock would be very harmless.

BY MR. MILLER:

Q How does Anthropic determine how to trade off helpfulness against harmlessness?

MS. MEHTA: Same objections. Objection to the form.

THE WITNESS: That is a very difficult problem that we face. I think one of our goals in the industry is to show that we actually can walk this line in an effective way. And I think certainly with Claude 3, when those models came out, people were impressed because instead of a blanket refusal for topics that would otherwise be sensitive, which a lot of other model providers were sort of exhibiting, the Claude models would

102

sort of try to explain the principles around those refusals, and explain the -- like, make the user not feel like they were doing something wrong for just asking if it was a well-meant query.

So I think despite the tension there, I think there's a lot that we can do to make things better.

BY MR. MILLER:

Q So when a user asks Claude for what the lyrics are to a particular song, the helpful response would be for Claude to output the lyrics to that song to the extent it knows them, right?

MS. MEHTA: Objection to the form.

THE WITNESS: I think it would somewhat depend on the user intent. But in the simplest case, yes, that sounds right.

BY MR. MILLER:

Q So absent some form of training or intervention that that output is harmful, a model trained to be helpful would attempt to provide lyrics in response to a prompt like that, right?

MS. MEHTA: Objection to the form. Incomplete hypothetical. Calls for speculation.

THE WITNESS: I guess in the Anthropic way of describing things, I don't think we would

103

consider copyright in any of helpful, honest, or harmless -- this is instead more of like a baseline of what we want our models to do or not do in the same way that -- never mind. I won't give an analogy. And instead it's more about our acceptable use policy and enforcing that.

BY MR. MILLER:

Q Do you recall earlier today we discussed memorization?

A Yes.

Q Does Anthropic use models to determine what content the models Anthropic is training should memorize?

MS. MEHTA: Objection to the form. Assumes facts not in evidence. Misstates the prior testimony.

THE WITNESS: [redacted]

BY MR. MILLER:

Q Why didn't it go anywhere?

MS. MEHTA: Form.

104

THE WITNESS: We do a lot of experiments that don't yield fruit, that don't show enough promise to continue investing in. When I say "a lot," I mean thousands. The project that I just described was sort of like a single afternoon kind of project. It didn't really work well enough. We moved on.

BY MR. MILLER:

Q I think you said earlier that you built the original content compare system; is that right?

MS. MEHTA: Objection to the form.

THE WITNESS: I did build the content compare system.

BY MR. MILLER:

Q Was that at one point called "Copy Fairy"?

A Yes, I named it Copy Fairy.

Q Where did the name come from?

A I came up with it, and it was kind of a reference to copyright fair use, portmanteau, with a little cutesy ending. But we've since renamed it.

Q When did you begin working on the Copy Fairy?

A The first time I prototyped it was over a

Pages 113-120

113

in native format, so if you scroll -- that's how it was produced.

MR. MILLER: Let's skip this one. We'll come back to it, then.

Go to tab 25.

(Plaintiffs' Exhibit 207, January 4/5, 2024, e-mail from Supriya Prabhakar to Sam Martin and Ben Mann, Subject: Re: Dataset for copyright filter checks, Bates stamped Anthropic_0000218341 through Anthropic_0000218342, was marked and presented for purposes of identification.)

BY MR. MILLER:

Q In the Zoom, you should see what's being marked as Plaintiffs' Exhibit 207, which, for the record, has the Bates number Anthropic_218341.

A I see that.

Q Okay. And does this appear -- excuse me. Does Exhibit 207 appear to be a continuation of the e-mail chain we were looking at in Exhibit 206?

MS. MEHTA: Objection to the form.

THE WITNESS: I'll take a minute to review this.

It looks like it is the same thread, but produced from the other side, since this e-mail

114

says -- oh, actually, this one is still to Sam. And now we know the full name of Rama P, and I think he went by Sandeep when I was working with him.

So, yes, it does appear to be the same thread.

BY MR. MILLER:

Q And the top e-mail on Exhibit 207 is from Supriya Prabhakar, whose name I apologize for probably butchering the pronunciation of, sent on January 5, 2024, to Mr. Martin and you, correct?

A Yes, that looks correct.

Q And Supriya Prabhakar writes, We completed the evaluation and observed a suppression rate of 26.22 percent.

Do you see that?

A Yes.

Q Do you know what "suppression rate" means in this context?

MS. MEHTA: Objection to the form. Foundation. Speculation.

THE WITNESS: I don't know.

MR. MILLER: Go to tab 27.

(Plaintiffs' Exhibit 208, January 8, 2024, e-mail from Evan Frondorf, Subject: Upcoming

115

ability to intervene on prompts, Bates stamped Anthropic_0000218344, was marked and presented for purposes of identification.)

BY MR. MILLER:

Q In the Zoom, you should see what has been marked as Plaintiffs' Exhibit 208, which bears the Bates number 218344.

A I see that.

Q And does Exhibit 208 appear to be an e-mail dated January 8, 2024, sent by Evan Frondorf?

MS. MEHTA: Objection. Lacks foundation. Calls for speculation. Form.

THE WITNESS: It does appear to be that kind of e-mail.

BY MR. MILLER:

Q Are you familiar with Evan Frondorf?

A Yes.

Q Who is Evan Frondorf?

A During our implementation of Claude on Amazon Bedrock, Evan was our trust and safety lead.

Q And did you work with Mr. Frondorf on the implementation of Claude on Amazon Bedrock?

A Yes, I did.

116

Q And this e-mail is sent to -- it appears Sophia Shramko and Harshal Pimpalkhute.

Do you see that?

MS. MEHTA: Objection to the form.

THE WITNESS: Yes.

BY MR. MILLER:

Q And do you know who those individuals are?

A Those are two people who worked at the Amazon on the Bedrock side at the time.

Q Directing your attention to the second paragraph which begins, [redacted]

Are you familiar with the [redacted]

MS. MEHTA: Form.

THE WITNESS: Yes.

BY MR. MILLER:

Q And what was the [redacted]

A [redacted]

117

[redacted]

Q Were the Anthropic models available on [redacted] the same as the Anthropic models available on Anthropic's first-party services?

MS. MEHTA: Objection to the form. Vague.

THE WITNESS: [redacted]

BY MR. MILLER:

Q Excuse me.

And then the first sentence of the second paragraph continues, [redacted]

Do you see that?

MS. MEHTA: Objection to the form.

THE WITNESS: Yes, I see that.

BY MR. MILLER:

118

Q So this feature at -- withdrawn.

Are you familiar with the feature being referred to here?

MS. MEHTA: Form.

THE WITNESS: I believe I am.

BY MR. MILLER:

Q And what do you understand that feature to be?

A I believe this is referring to our Prompt Shield technology.

Q So as of Monday, [redacted]

MS. MEHTA: Objection to the form. Assumes facts not in evidence.

THE WITNESS: Well, at the time, I believe [redacted] would have been the only [redacted] Prompt Shield [redacted].

BY MR. MILLER:

Q Why would a feature like Prompt Shield [redacted] first-party services [redacted]

119

MS. MEHTA: Objection to the form. Calls for speculation.

THE WITNESS: In general, and from my long experience as a software engineer, working between teams at a single organization is hard, but working across company boundaries is extremely difficult. There are many kinds of misalignments of incentives and interests, difficulties in sharing information that is basic and yet required to do the job, differing priorities on both sides for the different engineering teams, staffing issues, many other such things.

And this means that regardless of the feature, getting anything done in this complicated environment, especially with a hyperscaler like Amazon and a teeney-tiny start-up, which Anthropic was at the time, is extremely difficult. So I think the sort of baseline expectation for this kind of thing is that everything is extremely hard, let alone this fairly complicated and dynamic feature.

BY MR. MILLER:

Q When you were developing and implementing Copy Fairy, did you play any role in selecting the content to include in Copy Fairy's reference

120

corpus?

MS. MEHTA: I'm going to advise the witness just to be cautious with respect to any attorney-client communications or any work that you did at the direction of counsel. Insofar as you can answer that independent of attorney-client privilege or work done at the direction of counsel, then you can do so.

THE WITNESS: When I was prototyping the service, I selected the corpora that I would use. And I don't remember in particular what I used, but it didn't really matter, because we were just looking at things like the amount of data that would be hosted, the latency characteristics of the system, cost characteristics.

But subsequently, I can't provide any more details because I believe it would potentially divulge attorney-client privileged information.

BY MR. MILLER:

Q The question was did you play -- you -- withdrawn.

You're not a lawyer, right?

A That's correct.

Q So the question was, did you play any role in selecting the content to include in Copy

Pages 137-144

137

Do you have Exhibit 209 in front of you?
A Yes.
Q Do you recognize what's been marked as Exhibit 209?
A I'll skim the document.
Okay. I've reviewed the document.
Q Do you recognize what has been marked as Plaintiffs' Exhibit 209?
A I don't specifically recall this document, but it does look familiar, and I am familiar with the content.
Q Do you see at the top of this document, it states, DRI Miranda Zhang?
A Yes.
Q What does DRI mean?
A DRI is an acronym that stands for directly responsible individual.
Q Do you know who Miranda Zhang is?
A Miranda Zhang started at Anthropic as a recruiting coordinator. But since then, she has had a couple of different roles. I don't know what her role was at this time.
Q Did Anthropic hold an offsite meeting or retreat in June 2023?
MS. MEHTA: Objection to the form.

138

THE WITNESS: I don't specifically recall.
BY MR. MILLER:
Q Would you have attended any such offsite meeting that might have been held around that time?
A It's likely I would have attended, yes.
Q Do you recall attending an offsite meeting in late June 2023?
MS. MEHTA: Form. Asked and answered.
THE WITNESS: I don't specifically recall attending an offsite. But if it happened, I probably attended.
BY MR. MILLER:
Q Who else typically attends Anthropic offsites?
MS. MEHTA: Form.
THE WITNESS: Well, we do both team-specific offsites and also companywide offsites. This one doesn't seem to be specific to a team, so perhaps it's a companywide offsite. In which case, it would have been everybody at the company who could make it.
BY MR. MILLER:
Q In June 2023 -- withdrawn.
Do you recall how many employees,

139

approximately, Anthropic had in June 2023?
A Hmm. Not specifically.
Q Do you recall a ballpark?
MS. MEHTA: Objection to the form.
THE WITNESS: It probably would have been between a hundred and a thousand.
BY MR. MILLER:
Q Can you turn to page 23, please?
A Okay.
Q Do you see the last un-indented bullet on the page states, Where in the process does Constitutional AI or CAI come in? How do you explain the differences between models trained in different ways?
MS. MEHTA: Form.
THE WITNESS: Yes, I see that.
BY MR. MILLER:
Q And then the last sentence on this page states, For some very hard tasks, we can't use CAI.
Do you see that?
MS. MEHTA: Form.
THE WITNESS: Yes.
BY MR. MILLER:
Q Were there -- do -- withdrawn.

140

Do you know what that means?
MS. MEHTA: Objection to the form.
THE WITNESS: I could guess based on the context of the document, but I'd be speculating.
BY MR. MILLER:
Q Are there certain safety-related training objectives for which Constitutional AI is not effective?
MS. MEHTA: Objection to the form.
THE WITNESS: I guess independent of this document, I can tell you that there are many kinds of tasks for which Constitutional AI is not effective.
BY MR. MILLER:
Q What kinds of tasks -- withdrawn.
For what kinds of tasks is Constitutional AI not effective?
MS. MEHTA: Form.
THE WITNESS: So, in general, there are a couple of classes of tasks for which Constitutional AI is not effective. The biggest class of those is if the model doesn't already have some understanding of the problem at hand, then Constitutional AI will not help you, because the requisite knowledge is outside the model.

141

So to give an example of this, let's say that the model had never seen any text in the human-designed language Esperanto. In that case, there's no way to use Constitutional AI to make the model better at Esperanto, because Constitutional AI is a self-bootstrapping technique in which you ask the model to critique its own outputs and then improve its responses. But since that knowledge of Esperanto lives outside the model, then Constitutional AI will have no improvement.

BY MR. MILLER:

Q You said there were a couple classes of tasks for which Constitutional AI is not effective. What other types of tasks is Constitutional AI not effective on?

A Another one would be where it's a hard-skill problem. So let's take coding as an example. For a software engineering and programming in general, what you really need is practice, just sort of thinking about coding isn't going to necessarily make you better at software engineering.

And so this is another case where instead of -- the model already has a lot of coding and

142

programming concepts, but even though that knowledge is in the model, what it needs is a real environment in which to practice those skills. So that would be another class in which Constitutional AI would be unlikely to help.

Q Are there safety-related tasks for which Constitutional AI is ineffective?

MS. MEHTA: Objection. Form. Vague.

THE WITNESS: Yes, likely there are safety-related tasks.

BY MR. MILLER:

Q For what safety-related tasks is Constitutional AI ineffective?

MS. MEHTA: Same objections.

THE WITNESS: So one class of areas here is things that involve jailbreaks, so, typically, our Constitutional AI setup involves relatively short prompts and fairly non-adversarial inputs, whereas, jailbreaks are sort of by their nature meant to be out of distribution inputs so that force the model to do things that its trainers or creators didn't intend.

And so that would be another class of things that Constitutional AI would likely not help as much with, at least as it was implemented

143

in the paper.

BY MR. MILLER:

Q Okay. You referred in that answer to jailbreaks. What is a "jailbreak"?

MS. MEHTA: Objection to the form.

THE WITNESS: In a machine-learning context -- or, I guess, language model specifically -- a jailbreak is a specialized form of prompting that is designed to get a model to produce an output that with -- without the specialized prompting techniques, typically, that model wouldn't produce.

MS. MEHTA: Corey, we're a bit over. I think we're about five minutes past nine hours for today. If you could wrap up this line of questioning up --

MR. MILLER: Oh, I didn't even see -- you're right. Okay. We can go off the record, then.

THE VIDEOGRAPHER: Are we closing or just going off the record?

MR. MILLER: We can close for the day.

THE VIDEOGRAPHER: Okay. Please stand by. This marks the end of today's deposition of Benjamin Mann. We are going off the record at

144

20:34.

(Off-the-record discussion.)

MS. MEHTA: Can we just go on the transcript record? I want to note that the afternoon section of today is being designated highly confidential attorneys' eyes only under the protective order in this case.

MR. MILLER: Okay. And I will note my standard objection that we have no objection to the provisional designation under the protective order of the transcript as highly confidential attorneys' eyes only.

To the extent counsel is attempting to permanently designate the transcript as highly confidential attorneys' eyes only, we would object to that as inconsistent with the protective order.

(The deposition was adjourned at 8:35 p.m.)

* * *



Planet Depos®
We Make It Happen™

HIGHLY CONFIDENTIAL
ATTORNEYS' EYES ONLY

Transcript of Benjamin Mann, Volume 2

Date: November 21, 2025
Case: Concord Music Group -v- Anthropic

Planet Depos
Phone: 888.433.3767
Email: transcripts@planetdepos.com
www.planetdepos.com

146

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

- - - - - - - - - - - - - - - - x
CONCORD MUSIC GROUP, INC., et al., : Civil Action No. 5:24-cv-03811-EKL-SVK
Plaintiffs, :
v. :
ANTHROPIC PBC, :
Defendant. :
- - - - - - - - - - - - - - - - x

HIGHLY CONFIDENTIAL
ATTORNEYS' EYES ONLY

VIDEOCONFERENCED (HYBRID) AND VIDEO-RECORDED
DEPOSITION OF BENJAMIN MANN
PURSUANT TO FEDERAL RULE OF CIVIL PROCEDURE 30(b)(1)
San Francisco, California
Friday, November 21, 2025
12:07 p.m. Pacific Standard Time

VOLUME II

Job No. 609873
Pages: 146 - 226
Reported By: April L. Crites, RMR, CRR, CSR, CCR
California Certified Shorthand Reporter

147

DEPOSITION OF BENJAMIN MANN
HELD AT THE OFFICES OF:

WILMER CUTLER PICKERING HALE and DORR, LLP
50 California Street
Suite 3600
San Francisco, California 94111
(628) 235-1000

* * *

Pursuant to notice, and pursuant to the Federal Rules of Civil Procedure, before April L. Crites, RMR, CRR, CSR, CCR, and California Certified Shorthand Reporter No. 14493.

* * *

148

I N D E X

EXAMINATIONS CONDUCTED PAGE
BENJAMIN MANN
Cross-Examination by Mr. Miller..... 153

* * *

EXHIBITS

MARKED	DESCRIPTION	PAGE
Exhibit 210	March 13, 2024, e-mail from Ben Mann to nick@anthropic.com, Subject: [Auto summary] Ben/Nicholas (2024-03-12 14:03 GMT-7)-Transcript, Bates stamped Anthropic_0000218323	156
Exhibit 211	March 13, 2024, e-mail from Ben Mann, Subject: [Auto-summary] Data bottleneck (2024-03-12 15:34 GMT-7)-Transcript, Bates stamped Anthropic_0000012615	165
Exhibit 212	Jack Clark on X post: Technological Optimism and Appropriate Fear, Children in the Dark	173
Exhibit 213	Time Magazine article: Pausing AI Developments Isn't Enough. We need to shut it all down	178
Exhibit 214	Prompt, Bates stamped Anthropic_0000431956	214
Exhibit 215	Prompt: Stevie Wonder lyrics, Bates stamped Anthropic_0000524147	216

* * *

149

Pages 198-201

198

When it comes to the amount of the data that's coming in, in the case of books in our training process, we have a chopping process. That means the model will often not see a book in its original format but instead will be cut up into little pieces. For things like song lyrics, the same thing could happen, where a song gets chopped into pieces and the model never sees the whole thing, if it were present in training data at all.

Oh, I guess for purpose, I would also say transformativity. Language models, we have a lot of research that pushes the models to generalize from works -- that's the thing we really care about -- and away from memorization. We've published some of this work, and we go to great lengths to avoid memorization and favor generalization.

Jumping back to nature, as we prioritize getting various kinds of datasets -- you probably have seen from the evidence and exhibits that we've produced for this case that we heavily prioritize things that are factual. And, in particular, STEM is a big focus for us to try to achieve Dario's vision of machines of loving

199

grace, of improving health outcomes, education, emotional well-being, et cetera, for humanity.

And then, finally, on the effect on society, we've discussed at length our Societal Impacts team's work where we're doing our best to quantify the effects of our models on society.

In the case of music lyrics, the evidence that we have shows that it is a vanishingly small fraction of how our models are used today and the value that people get out of them. So it would be extremely surprising to me, although we don't have hard evidence, if there was a significant impact on, for example, the music industry because we used our -- because we used lyrics in our model training.

BY MR. MILLER:

Q All right. I want to ask you a little bit about what you said about the first factor.

And -- and you focused on that you expect the model outputs to -- to not be regurgitations of the underlying content.

Am I understanding you correctly that because the model is designed to output new content, you think that makes it fair use to use existing content in the training of the model?

200

MS. MEHTA: Objection to the form. Hold on one second. Objection to the form. Legal conclusion. And it misstates the testimony.

You can answer, but, again, same caution. Only answer as to your personal understanding, independent of advice of counsel.

THE WITNESS: I'll reiterate that I'm not a lawyer, and I'm not a copyright expert. And despite my interest in the subject, for what actually constitutes fair use, I would defer to people who spend their whole careers researching this topic.

So as far as our models' outputs, we have a multilayered system that I believe both I and Vinay Rao have spoken at length about how we attempt to prevent the regurgitation of our outputs from our training corpus, and elsewhere as well, of content that we believe the model shouldn't be -- shouldn't be outputting.

Of course, this is a constantly evolving cat-and-mouse game. We have baked training directly into our model to try to better solve this problem, and we believe we have one of the best systems at preventing unwanted regurgitation in the industry. But we're always looking to

201

improve it. And we take feedback from both our users and content owners in that process.

I'm not sure if that completely answers your question.

BY MR. MILLER:

Q It doesn't, so I will try to clarify it and ask it again.

So what we've been discussing is your views on the using of -- on the use of copyrighted data for training a model. And in that context, you've pointed to the fact, or what you claim is the fact, that the model -- that Anthropic's models are not intended to regurgitate copyrighted content.

And my question is, in terms of the fair -- of fair use for training, do you think whether something -- whether use of content for training is fair use depends on what the outputs of the model are?

MS. MEHTA: Objection. Misstates the prior testimony. Asked and answered. Calls for a legal conclusion.

And I give you the same caution: You can answer as to your personal understanding, independent of advice of counsel.