EXHIBIT 139

1

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

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

Highly Confidential
Attorneys' Eyes Only
Pursuant to the Protective Order

Videotaped Deposition of BEN ZHAO, Ph.D.
Chicago, Illinois
Friday, February 20, 2026
9:06 a.m. CST

Job No.: 621063
Pages: 1 - 462
Stenographically Reported By:
Michelle M. Yohler, CSR, RMR, CRR

2

Deposition of BEN ZHAO, Ph.D., held at the location of:

ARNOLD & PORTER
70 West Madison Street, Suite 4200
Chicago, Illinois

Pursuant to notice before Michelle M. Yohler, a Certified Shorthand Reporter, Registered Professional Reporter, Registered Merit Reporter, Certified Realtime Reporter, and a Notary Public, CSR No. 84-4531, in and for the State of Illinois.

3

A P P E A R A N C E S

ON BEHALF OF THE PLAINTIFFS:
MR. KEITH HOWELL
MR. TIMOTHY CHUNG (Remotely)
OPPENHEIM & ZEBRAK, LLP
4530 Wisconsin Avenue, NW, Fifth Floor
Washington, D.C. 20016
202.480.2999
khowell@oandzlaw.com

ON BEHALF OF THE DEFENDANT:
MS. SONAL N. MEHTA
WILMER HALE
2600 El Camino Real, Suite 400
Palo Alto, California 94306
650.858.6000
sonal.mehta@wilmerhale.com
-and-
MR. JARED V. GRUBOW (Remotely)
WILMER HALE
250 Greenwich Street
New York, New York 10007
212.230.8800
jared.grubow@wilmerhale.com
(Continued)

4

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

ALSO PRESENT:
Mr. Ian Chen,
In-House Counsel for Anthropic
Mr. James Porter, Videographer

5

E X A M I N A T I O N S

WITNESS	PAGE
BEN ZHAO, PH.D.	
By Ms. Mehta	7
By Mr. Howell	425
By Ms. Mehta	428

E X H I B I T S

EXHIBITS		PAGE
No. 1	Expert Report of Dr. Ben Zhao Dated December 19, 2025	172
No. 2	Rebuttal Expert Report of Dr. Ben Zhao Dated January 23, 2026	172
No. 3	"somesite I used to Crawl: Awareness, Agency and Efficacy in Protecting Content Creators From AI Crawlers."	258
No. 4	Reply Expert Report of Dr. Ben Zhao Dated February 13, 2026	318

6

P R O C E E D I N G S

THE VIDEOGRAPHER: Here begins Media Unit 1 in the video deposition of Ben Zhao in the matter of Concord Music Group Inc., et al., versus Anthropic PBC in the United States District Court, Northern District of California, San Jose Division, Case Number 5:24-cv-03811-EKL-SVK.

Today's date is February 20th, 2026, and the time on the video monitor is 9:06 a.m.

The videographer today is James Porter representing Planet Depos.

This video deposition is taking place at Arnold & Porter in Chicago, Illinois.

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

MS. MEHTA: Good morning. Sonal Mehta from WilmerHale on behalf of Anthropic.

MR. HOWELL: Good morning. Keith Howell from Oppenheim & Zebrak on behalf of the plaintiffs.

THE VIDEOGRAPHER: And do we have counsel on Zoom?

MR. GRUBOW: Jared Grubow from WilmerHale on behalf of Anthropic.

MR. CHUNG: Good morning. Timothy Chung on behalf of Publishers.

MS. MEHTA: And then I believe we have in-house counsel from Anthropic also on the line.

THE VIDEOGRAPHER: The court reporter today is Michelle Yohler representing Planet Depos.

Please swear in the witness.

(WHEREUPON, the witness was duly sworn.)

BEN ZHAO, Ph.D.,

called as a witness herein, having been first duly sworn, was examined and testified as follows:

EXAMINATION

BY MS. MEHTA:

Q Good morning.

A Good morning.

Q Dr. Zhao, could you please state your first -- or full name and your address for the record, please.

A Sure. Ben Zhao, my -- I'm sorry. You said my address?

Q Yes.

A 130 North Garland Court, Chicago, Illinois, 60602.

Q What do you understand this case to be about?

8

MR. HOWELL: Objection. Form.

BY THE WITNESS:

A I understand this case is about Anthropic's training of data and Anthropic's design of its large language model and its mechanism to operate that model.

BY MS. MEHTA:

Q What legal claims do you understand to actually be at issue in the case?

MR. HOWELL: Objection. Form. Calls for a legal conclusion.

BY THE WITNESS:

A I am not here to speak to the legality of the claims; I'm here as a technical expert, so, as such, I speak to the technical components of the case.

BY MS. MEHTA:

Q Understood. But you understand that in courts of law, the people actually make claims based on legal challenges and then defend on legal challenges --

A Yes.

Q -- right?

A Yes.

MR. HOWELL: Objection. Form.

149

BY MS. MEHTA:
Q Would you have the ability to design an experiment to test the guardrails and the efficacy of the guardrails if you so choose?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A Would I be able to? Yes, of course. I mean, I don't see a particular need to. There's a lot of published reports that give very specific guidance on how to form, you know, what are almost benchmarks for these types of guardrails. And there's well-known methodology that, you know, researchers are taking when they assess the capabilities of these kind of guardrails.
So I don't see a need to reinvent the wheel, but, you know, many of the methodologies are fairly familiar to me. They're, you know, common knowledge in the research literature.
So, yeah, I don't see a need to invent new ones, but, you know, the existing methodologies certainly all make sense and it's things that I understand fairly well, yes.
BY MS. MEHTA:
Q Right. So the question I was getting at is -- I'll ask it a slightly different way.

150

Was there anything stopping you from conducting your own test of the efficacy of Anthropic's guardrails post 2023?
MR. HOWELL: Objection. Form. Vague. Calls for speculation. Asked and answered.
BY THE WITNESS:
A I believe there were higher priority things that demanded my time. There is plenty of research from people that I know and respect in the field who have done fairly exhaustive tests who have demonstrated the -- the, sort of, ineffectiveness of some of these guardrails.
And so, no, I did not see it necessary to repeat their tests because I honestly, you know, have higher priorities.
BY MS. MEHTA:
Q Did you -- you mentioned the -- a number of people in the field that you respect that engaged in these tests. You understand that a number of the tests have been characterized or criticized as involving adversarial prompting, right?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Not sure "criticize" is the right word.

151

When we ascertain the efficacy of any system, you don't test it with all the most naive and basic conditions because then you don't learn anything.
So, you know, whether it's computer security or safety or any of these fields, the common practice is to understand where the limitations of systems are, and if you just give them the most common and basic test, that's not going to accomplish the purpose.
BY MS. MEHTA:
Q Are you aware of any test of the efficacy of Anthropic's guardrails with respect to copyright regurgitation that did not involve adversarial prompting or jailbreaking?
MR. HOWELL: Objection. Form. Vague. Compound.
BY THE WITNESS:
A I believe Dr. Bilmes in his report mentioned his own tests, which seemed to be largely non-adversarial.
But certainly I believe in most of these reports that are out there, the reports tend to cover a range of prompts starting from the most naive and basic to techniques that are increasingly specialized or, you know,

152

experimental.
BY MS. MEHTA:
Q Setting aside Dr. Bilmes' opinions, which I know you disagreed with. We'll get into that later.
With respect to the academic literature and the research that you relied upon for your opinions in this case, are you aware of any research or study that has tested Anthropic's guardrails with respect to copyright regurgitation and successfully elicited copyrighted material without jailbreaking or using adversarial prompting?
MR. HOWELL: Objection. Form. Vague. Asked and answered.
BY THE WITNESS:
A That's a very specific question. I mean, if I had time, I could go back and revisit the conditions for a bunch of these research reports and answer that more concretely.
Offhand I would say probably because I remember especially some of these older reports, you know, when people were first starting to look at both hallucinations and memorization, you know, many of the initial tests were more

153

non-adversarial, and I believe that at least some of them produced results.

And that's really what prompted much of the research that followed up later on when people realized that memorization and regurgitation were issues, and then they did more complete tests that covered even wider range of prompts.

But, you know, as far as, like -- you know, in addition to these types of reports, I mean, my own experiments back in 2023, many of them were -- of my requests were non-adversarial, and they still produced regurgitations that were verbatim.

Q Let me be specific as to time and as to the particular models that I'm asking about. I'm not asking about ChatGPT or Gemini or anyone else's models; I'm asking specifically about Anthropic's commercial models, and I'm asking specifically about 2024 -- post 2024, so 2024, 2025, and '26.

Are you aware of any study, other than Dr. Bilmes' study in this litigation, any study in which someone has prompted Claude with respect to copyright -- let me actually take that back. Let me start again.

154

In 2024, 2025, or 2026, are you aware of any study testing the efficacy of Claude's guardrails with respect to copyright regurgitations in which they were able to elicit copyright material without adversarial prompting or jailbreaking?

MR. HOWELL: Objection. Form. Vague.

BY THE WITNESS:

A I mean, you're asking me a question as if I'm a literature database. I wish I could have all the details at hand in my mind right now.

I will just say that a number of them have produced results that were significant regurgitations. You know, some of these papers covered entire ranges of different queries from the benign to the adversarial.

And so do I know the exact nature of the queries that produced regurgitations? Not off the top of my head, no. I can -- you know, I would need to go back to the literature and really zone in on the specific results in certain tables, exactly the -- which questions were being prompted and which produced some of these results.

You know, so -- so I can't answer at that level of detail. I will just say that I do recall

155

at least some paper abstracts talking about, at a high level, how easy it was to solicit regurgitations from models including Anthropic, including Anthropic models that are more recent, you know, thinking of, you know, Claude 3-point-something and -- and above.

BY MS. MEHTA:

Q Did you -- we're going to take a break in a second, I just have one follow-up question to that which is: Did you -- I don't remember seeing anywhere in your reports an identification of any copyright regurgitation in a test of Anthropic's copyright guardrails where there was copyrighted material elicited in the absence of adversarial prompting or jailbreaking.

It's fair to say that you have not identified that in your reports, correct?

MR. HOWELL: Objection. Form. Vague. And to the extent you're asking him about a document without showing it to him.

BY THE WITNESS:

A I have identified citations to reports that have talked about regurgitations from Anthropic, and in more recent times covering the models that are, you know, more recent. I don't

156

believe I explicitly noted whether they produced results with benign prompts or not, but those reports were certainly cited.

I just don't think I commented on the exact nature of the prompts, so, to that question, there was not an explicit note about benign prompts presenting them, but there were certainly multiple citations to reports that talked about modern versions of Anthropic models or more recent versions of Anthropic models regurgitating contents -- or regurgitating --

BY MS. MEHTA:

Q Right.

A -- content.

Q What I was getting at was just said, which is: You didn't go through and analyze for purposes of your opinions in this case which specific prompts generated regurgitation in the literature to determine whether those prompts were benign or ordinary or adversarial or jailbreak, fair?

MR. HOWELL: Objection. Form. Vague. Asked and answered.

BY THE WITNESS:

A And, as I will repeat, most of these tests

157

span the gamut of benign to adversarial. As a baseline, typically studies of this type always provide some, you know, initial starting point.

You know, when you do a scientific study, you need a baseline; you don't just say I'm going to test exactly this and nothing else. So, generally speaking, these reports cover quite a range of different approaches and then, you know, come to high-level conclusions.

So, no, I did not explicitly mark down, you know, the conditions and the prompts for some of these queries and exactly which prompts of which type solicited how much regurgitation.

MR. HOWELL: All right. Let's take our break.

THE VIDEOGRAPHER: Going off the record. 11:48 a.m.

(WHEREUPON, a recess was had.)

THE VIDEOGRAPHER: This begins Media Unit 3. We are back on the record at 11:58 a.m.

BY MS. MEHTA:

Q Dr. Zhao, for purposes of your opinions in this case, have you done any analysis, study, evaluation, of the typical ways in which users prompt Claude as it relates to music or song

158

lyrics?

MR. HOWELL: Objection. Form. Vague.

BY THE WITNESS:

A I think you're asking if I've done a user study to ascertain how common people or common users interact with lyrics? No.

BY MS. MEHTA:

Q Aside from a user study, have you done anything to evaluate or analyze or study how common users use Claude?

MR. HOWELL: Objection. Form.

BY THE WITNESS:

A I mean, a user study would be the appropriate way to do that, which I have not conducted, so no.

BY MS. MEHTA:

Q Okay. You would agree with me that Claude does not maintain or store complete copies of the Publishers' works in suit in its training data?

MR. HOWELL: Objection. Form.

Go ahead.

BY THE WITNESS:

A There's a lot of nuances there, but I'm pretty sure that I would not agree to that statement.

159

BY MS. MEHTA:

Q Well, would you agree with the following: After training Claude, the system does not store passages of text, or in the case of the Publishers' works, the song lyrics, from its training data?

MR. HOWELL: Objection. Form. Misstates prior testimony. Asked and answered.

BY THE WITNESS:

A So, again, you know, these are the issues that we're talking about, right? And, you know, depending on your definition of "store" and there's a lot of technical issues here, I would not make that statement.

BY MS. MEHTA:

Q You opine in your report that Claude maintains a statistical representation of training data in its systems.

MR. HOWELL: Objection. Form. Same objection regarding documents not in front of him.

BY THE WITNESS:

A I've made a number of statements that are relevant to what is stored inside the Claude model. I would say statistical representation is a little bit of an oversimplification.

160

Yeah, I would just leave it at that. So I would not agree to that particular assessment.

BY MS. MEHTA:

Q Okay. Would you agree to the sentiment that, with respect to LLMs generally, they are probabilistic and not deterministic?

MR. HOWELL: Objection. Form. Vague.

BY THE WITNESS:

A What I will say is that computers are deterministic until they are made stochastic.

And these models, to the extent that they are deterministic software programs sitting on a computer, without explicit parameters to increase stochasticity, things like the temperature parameter, they would be deterministic and they would give the same result every time.

That is the nature of physics and computers in general, determinism.

BY MS. MEHTA:

Q I'm not asking about the computer; I'm asking about large language models.

A Large language models run on computers as software, and so my statement stands.

Software systems on computers do not randomly behave in unexpected ways unless there is

173

BY MS. MEHTA:
Q And, Dr. Zhao, if you could turn to the rebuttal report to start. That's Exhibit Number 2. And go to Paragraph Number 10.
Paragraph Number 10, you're responding to some statements from Dr. Bilmes in his report.
You see that?
A Yes.
Q By the way, did you know Dr. Bilmes before this case?
A Did not.
Q Had you ever looked at any of his research before this case?
A I did not. He is not in computer science; he is a professor of electrical and computer engineering, and so we do not cross paths that regularly.
Q So your position is or your view is that people that are in computer science don't regularly cross paths with people in electrical engineering?
MR. HOWELL: Objection. Form. Misstates prior testimony.
BY THE WITNESS:
A I'm in no way making such a

174

generalization.
BY MS. MEHTA:
Q Okay. Now, with respect to Dr. Bilmes, I take it then you haven't evaluated the quality of his academic research or writings, fair?
A It is not my opinion -- I do not give an opinion here about the quality of his research.
Q And you haven't formed such an opinion?
A I'm sorry?
Q You have not formed such an opinion?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I don't give such an opinion in this declaration or in any of these reports.
BY MS. MEHTA:
Q So I guess what I'm getting at is: You don't -- you don't intend to offer an opinion in this case to the judge or the jury about the quality of Dr. Bilmes' academic research, fair?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Not in general terms. I certainly engage and I interpret his comments and his statements with respect to relevant work and in these specific technical issues, yes.

175

BY MS. MEHTA:
Q At Paragraph Number 10 you say: "Dr. Bilmes is correct that a pre-trained model does not specifically recall and then display texts from a database of training corpora in response to user prompts."
Do you see that?
A Yes.
Q Do you stand by that statement?
A I stand by the statement because the keyword there is "from a database of training corpora."
Q And that's because the model does not pull from a database of training corpora when it is delivering responses to a user prompt?
MR. HOWELL: Objection --
BY MS. MEHTA:
Q Correct?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A That is because there is not a database inside these LLMs.
BY MS. MEHTA:
Q In fact -- and you can go to Paragraph 21 if this is helpful. You say in Paragraph 21 that

176

LLMs do not store passages of text in a central database, correct?
A Again, the keyword here is "central database." By "database," I mean a traditional definition of a computer database. Common examples include things like SQL server, MySQL, you know, Oracle database, maybe they've renamed it now.
But those types of databases, yeah.
Q And so using the common computer science dictionary -- or computer science definition of "database," LLMs do not store data or text in a database and retrieve that data, correct?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A By that very strict definition of database, that is true, but I will just remark that computers in general, very few computers actually rely on internal database. In fact, almost every piece of software that we use today does not include a formal database internally.
That does not mean that it does not store content. That does not mean that it does not transform and store different versions of the content in different representative forms.

177

You know, yeah, the idea that a strict definition of a database, yes, I agree that the LLMs do not store a strict definition of a database, but they are accessing different means and different representations.
BY MS. MEHTA:
Q Okay. I understand your opinions with respect to accessing the information; I'm asking specific questions with specific language intentionally.
You'd agree with me that LLMs are not data retrieval systems?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Depending on who you speak to, the definition of a data retrieval system is incredibly expensive, and so it would really depend on what you define to be a data retrieval system.
There are definitions of data retrieval systems where every single computer, regardless of whatever software they're running, are data retrieval systems. So, no, I would not agree with that statement.

178

BY MS. MEHTA:
Q Would you consider an LLM to be designed to -- actually, let me ask the question differently. Why don't you go to Paragraph 11 in the same report, the rebuttal report.
You say there: "The model has not memorized the specific text in the sense of literally storing an exact copy of that text and it does not strictly refer back to and fetch that sequence from its training corpus when it regurgitates. A pre-trained model does not scan for text from its training corpus and reproduce what it finds when asked."
Do you see that?
A Yes.
Q Do you stand by that statement?
A Yes, the model does not scan for the text from its training corpus and produce it exactly.
Q And the model does not -- when it's prompted by a user, it doesn't go to its training corpus, retrieve text, and then reproduce that text verbatim in response to a prompt, correct?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Can you restate the question?

179

BY MS. MEHTA:
Q Yes. When a user prompts a model like Claude, the model does not go to its training corpus, retrieve that text, and then reproduce or deliver that text to the -- to the user, correct?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A That is correct in the sense that I believe the training corpus is decoupled from the model once the model is trained.
BY MS. MEHTA:
Q And when you say "decoupled," you mean that the -- once the model is trained, it does not have access to the original training corpus, correct?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A It really depends on your definition of the original training corpus. Under the interpretation that the model parameters are, in fact, a lossy compressed version of the training corpus then, you know, it would have access, just in a slightly different format, a different representation format.

180

BY MS. MEHTA:
Q Well, I'm using the language you used. You said that once the model is trained, it's decoupled from the training corpus. And what you meant by that is --
A I meant by that the original copy of a training corpus.
Q Right. So the text, in the case of a text-based LLM like Claude, the text that is used to train the model, the model does not have access to that text when it -- after it is trained?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A No, I meant the actual original copy of the training corpus, that is the part that is decoupled. There -- you can make copies, you can make different representation copies of that training data and still maintain access to those copies.
BY MS. MEHTA:
Q In the case of Claude, Claude does not maintain a -- does not have access to or continue to access either the original training corpus or a copy of the training data after it's trained, correct?

273

Q And then there's a citation there to the deposition testimony of Benjamin Mann, right?
A Yes.
Q Do you -- other than Mr. Mann's testimony, do you know the -- I don't want you to repeat what he said. What I'm asking you is do you have any other knowledge about whether and to what extent Anthropic ever failed to abide by robots.txt?
MR. HOWELL: Objection. Form. Vague.
Go ahead.
BY THE WITNESS:
A So I can clarify. You're asking me, in addition to the multiple citations I have here in this particular paragraph, do I have more citations that restate the same thing.
And the answer is, no, that I have a citation to a business insider information with a full URL, and then I have a citation to Mr. Mann's own testimony. So you have both a public press report and testimony from an Anthropic employee.
You're asking me, in addition to public press and testimony -- sworn testimony from an Anthropic employee whether I have additional other types of information stating the same fact?

274

BY MS. MEHTA:
Q Supporting your opinion, that's what I'm asking you.
MR. HOWELL: Same objection.
BY THE WITNESS:
A No, it is only the testimony -- sworn testimony of an Anthropic employee and published public media articles.
BY MS. MEHTA:
Q Are you aware of any evidence anywhere in the record, including in the deposition testimony that you cited, that Anthropic ever scraped any of the works in suit or any song lyrics without abiding by robots.txt?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I'm sorry. Can you repeat the question again?
BY MS. MEHTA:
Q Yes. Are you aware of any evidence that Anthropic ever scraped lyrics for any songs, or for the works in suit in particular, without respecting robots.txt?
MR. HOWELL: Objection. Form.

275

BY THE WITNESS:
A So let me unpack that question for a second.
Am I aware -- do I have specific knowledge that Anthropic explicitly ignored robots.txt in scraping lyric content for its training data? The answer to that question is no.
However, Anthropic has gotten training data from a number of sources, not all of which have respected robots.txt; so web archives, YouTube video transcriptions, Common Crawl in the past.
So all of these things are things that do not require Anthropic to interact with robots.txt. Even though someone else may have gotten data on their behalf or relayed them -- data to them through some means and if they did not respect robots.txt to get that data, Anthropic would not have necessitated -- would not have been necessary for Anthropic to personally and directly ignore robots.txt, and, yet, they would still have been privy to data that was obtained through that process.
BY MS. MEHTA:
Q Can you identify the lyrics for any

276

specific works in suit where Anthropic obtained that data through -- either directly or indirectly in violation of or ignoring a robots.txt preference?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A So I would need to confirm this to double-check, but it is my belief that, at some prior point, Common Crawl did not always respect robots.txt and, therefore, data that was gathered by Common Crawl could have come from scrapers that ignored robots.txt.
And, again, if that data is fed into the training pipe, as I think there is evidence to show that Anthropic did, in fact, train on Common Crawl data, then that data would have been obtained potentially in violation of robots.txt.
BY MS. MEHTA:
Q Now, you understand that song lyrics are widely available on the internet through a number of different sites, right?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A I understand that they are available in different forms, different places. I also

277

understand that many of these places have actual licensed access to them and agreements with some of the plaintiffs and other music and music industry companies.

BY MS. MEHTA:

Q Are you aware of any website that included the lyrics to the works in suit that was scraped either directly by Anthropic or by one of these other data acquisition sources like Common Crawl in which the person that scraped -- or the crawler that scraped the website ignored robots.txt to obtain lyrics for the works in suit in this case?

MR. HOWELL: Objection. Form. Asked and answered.

BY THE WITNESS:

A I believe I've literally just answered this question with Common Crawl because, again, I believe that, at a prior time, Common Crawl did, in fact, ignore robots.txt.

And if that were the case, then there would be scenarios where Common Crawl would have scraped and downloaded content from any number of websites, some of which may have contained lyrics content.

And that content, when served up and

278

trained into the Anthropic pipeline would have satisfied the constraint that you're asking for.

BY MS. MEHTA:

Q That was not the question I was asking.

I was asking about a lyric website where they had a do-not-scrape preference turned on and where Anthropic either directly or through Common Crawl or one of these other sources scraped the lyrics. Can you name any such website?

MR. HOWELL: Objection. Form --

BY MS. MEHTA:

Q For any of the works in suit.

MR. HOWELL: Objection. Form. Asked and answered.

I'll just note the time as well. We're over an hour.

Go ahead.

BY THE WITNESS:

A So you're asking me to provide proof that, for a particular type of behavior for a particular subclass of websites, I will again, just -- all I can offer is that, to my knowledge, Common Crawl did not always respect robots.txt.

And if at any given time when it was not respecting robots.txt, any one of the number of

279

lyrics websites that we've listed in the reports -- and there's been a number of them, I don't remember what page, but they are there -- any of them, if they had robots.txt, would have been scraped and data gathered against their robots.txt and --

BY MS. MEHTA:

Q Right. And my question is not about "if." My question is: Are you aware of any evidence that any of those websites had robots.txt turned off or a preference against scraping set at the time that they were scraped? Any website with any of these songs on it.

MR. HOWELL: Objection. Form. Vague. Asked and answered.

BY THE WITNESS:

A I don't have direct access to Anthropic's website -- or, sorry, to Anthropic's web crawls of their datasets, and I'm not aware of whether Anthropic actually downloaded robots.txt or fetched it in the process.

I'm also not aware of whether Common Crawl, when it did or did not -- irrespective of whether it did or did not respect robots.txt, I'm not aware of whether they kept copies of those

280

robots.txt around for their own archival purposes or stored together with their own dataset, and so I can't possibly answer that question.

BY MS. MEHTA:

Q Do you -- are you aware of or have you done anything to look at any of the lyric websites that you listed in your reports to see whether, at any point in time, they had a do-not-scrape setting?

MR. HOWELL: Objection. Form.

BY THE WITNESS:

A If I can just sort of unpack your question. You're asking me, for all time, whether I'm aware of specific time zones and time periods during which any of these websites what their robots.txt policy is for.

And my answer is, no, I do not have a 24/7 multi-year tracker that has monitored the robots.txt content for all these websites, and, therefore, I could not tell you if, at any given time, there was a lapse or they changed their robots.txt or the contents of the robots.txt from any specific date to any other specific date.

BY MS. MEHTA:

Q Can you identify any instance, any

281

instance, any moment in time for any of the lyric websites where the lyric website had its robots.txt set to not scrape and they were scraped either by Anthropic or by someone elsewhere Anthropic then later acquired the data?
MR. HOWELL: Objection --
BY MS. MEHTA:
Q A single one.
A Again --
MR. HOWELL: Objection. Form. Asked and answered.
BY THE WITNESS:
A Asked and answered. And in order for me to tell you if such a thing occurred, I repeat: I would have had to have a constant set of crawlers and scrapers that constantly monitored these websites on a 24/7 basis frequently fetching, comparing, storing, all the robots.txt for all time since this suit started.
And I'm saying that that is not a reasonable question to ask, and, therefore, I do not know the answer.
BY MS. MEHTA:
Q No, sir, that's not what I'm asking. I'm not saying can you go through and comprehensively

282

tell me for every moment in time for years.
I'm asking whether you're aware of a single piece of evidence in this case that any lyric website had their robots.txt turned off to do not scrape?
A And I'm --
MR. HOWELL: Objection --
BY THE WITNESS:
A -- telling you --
MR. HOWELL: Same objection.
BY THE WITNESS:
A -- way to answer that question would had to have been through deployment of what I just described as a 24/7 archival and checking and -- frequently checking and fetching process by which I'm monitoring all these websites.
You're asking me a question that can only be answered if I had deployed such a mechanism, and my answer is, no, I did not deploy such a mechanism and, therefore, it is impossible to give you an answer to that question.
BY MS. MEHTA:
Q Have you done any investigation in -- either by asking counsel for data in this case or through research on the internet or in the

283

Internet Archive to determine whether there's any lyric aggregating website that had robots.txt turned off at any point in time where it was scraped either directly by Anthropic or by another scraper that Anthropic later acquired the data to?
MR. HOWELL: Objection. Form. Compound.
BY THE WITNESS:
A That was, like, a paragraph of questions. If you could break that down, that would be more helpful.
BY MS. MEHTA:
Q Yeah, it was one question. I'm just asking whether you have investigated -- did you ask counsel for or did you go onto the Internet Archive or do anything else to investigate whether there were any lyric-aggregating websites that had robots.txt turned off at any point in time?
I'll stop there.
MR. HOWELL: Same objections.
BY THE WITNESS:
A So --
MR. HOWELL: -- asked and answered.
BY THE WITNESS:
A -- by nature of this question, I just want to point out that you're asking questions that

284

require full examination of a time period. Sounds like you're saying "give me any time," but what that really means is study the entire time period to detect if there is any time for what particular condition or set of constraints has been met.
I will say that I have visited some of these lyric websites. I have, on occasion, I believe for at least one of them, looked at their robots.txt, okay?
However, that is insufficient to answer your question because, again, your question of "Do you know of any particular time certain constraints were met," which implies that if I say my one test or a handful of tests did not show this condition, then the implication is that that event never happened, which of course would be misleading.
So, again, I have, on occasion, at least in one specific case, I remember going to a lyrics website, looking at the format, looking at robots.txt, okay? But, again, in order for me to answer your question definitively or even clearly, I would have had to do this highly frequently over a long period of time and done an exhaustive comparison of how or if the robots.txt had changed

289

before or after cureration and, therefore, I cannot answer that question.
BY MS. MEHTA:
Q And when you say you had no direct access, you mean you did not go to Anthropic to inspect the data that was made available?
MR. HOWELL: Objection. Form.
BY MS. MEHTA:
Q Correct?
A That is correct.
Q Okay. Now, at Paragraph 61, you say that the tool was used during the scraping process to extract text from websites without it -- without what it referred to as, quote, useless junk contained in webpage footer -- headers and footers.
Do you see that?
A Yes.
Q Do you have an understanding as to whether Anthropic ever used any extractor tool to remove anything from the HTML body --
MR. HOWELL: Objection --
BY MS. MEHTA:
Q -- of a website that was scraped.
MR. HOWELL: Objection. Form.

290

BY THE WITNESS:
A So I believe my answer is in the report, and I will read: "Based on my review of the internal Anthropic discussions, it is my view that Anthropic utilized such extractor tools because they removed, quote/unquote, undesirable information and repetitive information such as licenses and other so-called garbage from scraped content."
And there's a citation there with a specific document number.
BY MS. MEHTA:
Q Okay. That wasn't an answer to my question, though.
My question was: Do you have an understanding as to whether Anthropic used any extractor tool to remove text from the HTML body as opposed to a header and a footer from a webpage?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A Header and footers are in the HTML body, so if you could clarify exactly why those two questions are consistent, I will --

291

BY MS. MEHTA:
Q How about --
A -- happily --
Q -- the question differently?
Have -- do you have any knowledge of or any evidence of Anthropic ever using the extractor -- an extractor tool to remove text from outside of a webpage header and footer?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I'm sure -- I'm not sure I know what you mean. So when you say outside of a webpage's header or footer, what does that mean?
BY MS. MEHTA:
Q Anything other than the header or footer?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A Oh, you mean other than -- not, like, physically outside, but other than, that's the --
BY MS. MEHTA:
Q Correct.
A -- the -- oh, I see.
I recall there's discussions of copyright content and other types of data that's been removed through the processing of data after it's

292

been collected, but I believe that was in Dr. Shan's report.
So, off the top of my head, I don't recall the details, but it's -- if I recall correctly, he discusses evidence of Anthropic removing copyright notices from content.
Yeah, and my impression, if I recall -- again, I don't recall the constraints or qualify -- qualifiers regarding headers or footers, so I believe that was referring to overall removal of copyright data.
And that would be an answer to the question, if I understand correctly, that -- that you're asking right now.
BY MS. MEHTA:
Q Okay. Thanks for that clarification about what you were referring to from Dr. Shan's report. I want to -- we'll have a chance to talk to Dr. Shan ourselves directly, so I want to focus on what you did yourself as opposed to the work that you read that he did.
So from your own work, are you aware of any instance in which Anthropic removed or extracted any text outside of a webpage header or footer?

293

MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A So everything that I know of regarding that subject is here in my report.
As to what exactly Anthropic may or may not have done internally after getting the data, again, you know, I think that would have been something that is internal to Anthropic. I'm not privy to that data.
I will also add that, based on my understanding of what was presented for inspection at the data inspection process, based on my understanding of what that data was, that that answer could also have not been answered by Dr. Shan as well because my understanding was that one version -- or the primary version of the dataset that he was privy to was pre-cureration raw data files.
Again, I could be off on some of the constraints and the qualifiers, but that is my high-level understanding.
And then the other version of that which involved, sort of, post-cureration data, you know, was presented in only tokenized form without a token table -- token translation table, and,

294

therefore, it was impossible for us to determine to what extent specific types of data survived the cureration process or what kind of data was removed in the process simply because we could not translate that data without the help of the tokenizer table, which we were -- we did not have access to.
BY MS. MEHTA:
Q Now, your understanding about what was and was not available in the inspection environment, that all comes from Dr. Shan, correct? Because you didn't go there and inspect it yourself?
A That's correct; I did not attend the data inspection myself.
Q Okay. So now going back to the question of extracting or removing text. Again, I asked the question specifically as to what you did as opposed to what Dr. Shan did.
Did you yourself do any analysis of any instance in which copyright management information was extracted or removed by Anthropic related to any of the works in suit?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Pretty sure we've answered this question,

295

but I will attempt to do so again.
Asking me whether I'm aware of Anthropic modifying content removing -- for example, removing things like copyright headers or footers or copyright notices or copyright information would have required direct access to post-cureration data, which not only did I not have access to, but I also believe Dr. Shan did not have access to because we did not see any version of post-cureration data that was interpretable or that could be compared to the original training data that was gathered and scraped.
BY MS. MEHTA:
Q Does that answer apply only to the works in suit, or would that apply to all training data?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I would not generalize my statements based on what I've already said, particularly since Dr. Shan may be privy to more information than I'm fully aware of, but, as far as I know, these statements are true for the data that we're talking about.

296

BY MS. MEHTA:
Q Okay. So let me be specific then when you say "the data that we're talking about."
With respect to the works in suit, you yourself have not analyzed whether Anthropic removed or extracted any copyright management information, correct?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A I'm sorry. Could you rephrase -- repeat the question?
BY MS. MEHTA:
Q Yes. I want you -- you said something about relevant data, so I just want to be very specific.
I'm talking about the works in suit, the 499 works in suit in this case. You yourself have not analyzed whether Anthropic removed or extracted any copyright information -- management information for the works in suit, correct?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Yeah, I do think that question is quite vague. I will say that -- more concretely, that I have not seen data that proves to me what

305

guardrails for preventing copyright regurgitation is complicated?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A Complicated? That's a -- again, it's -- there's just a lot of nuances here, so I will unpack a little bit.
You know, is there a single trivial thing that will solve all universal queries? Probably not. Is it reasonable to do a good job, is it possible to do a fairly good job and make it difficult for whatever undesirable content may be, including copyrighted content to be blocked or filtered? I would think it's possible. Of course to differing degrees.
And there's lots of different potential methodologies. You know, and they will be different in their technique but also in resources consumed. And, as you say, for example, your, you know, use of the word "complicated," you know, most of these software systems are reasonably complicated, so, you know, nothing as -- is something that someone's going to pop into a system, you know, after 30 minutes or an hour, so I guess complicated is all relative.

306

But, again, these are companies that have a ton of resources, not only in terms of compute and processing and data, but also in terms of highly qualified engineers who are able to handle this level of complexity.
BY MS. MEHTA:
Q You'd agree there are trade-offs associated with guardrails against copyright regurgitation?
MR. HOWELL: Objection. Form. Vague. Calls for speculation.
BY THE WITNESS:
A I mean, in general, all systems have trade-offs. In fact, you know, entire research areas in computer systems are, you know, characterized by trade-offs --
Q But in this --
A -- and --
Q Oh, I'm sorry.
A Go ahead.
Q In this particular instance, you'd agree there's a trade-off between having very tight filters to protect against copyright regurgitation, and the risk of false positives that will block content that is not copyright

307

regurgitation and that would be a valid or positive use case? You'd agree with that, right?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A There are always trade-offs. And of course the question is really priorities, right? And so, you know, if you think about the universe of different kind of queries that any one of us can ask an LLM or interact with or give it as prompts, I mean, there's a potentially unlimited universe of that type of questions, right?
And obviously Anthropic and other companies are in the business of expanding that, you know, on a daily basis.
So, you know, with respect to our -- you know, the entire universe of kind of things that LLMs can possibly respond to or not, you know, is it possible that some set of functionality might get affected by protective measures? Almost certainly.
But the question is not whether anything will be affected; the question is really, you know, again, as with many other cases in real systems, there are no absolutes and it's a question of what will get affected and how

308

important that is. And, again, how important is obeying the law and respecting copyright. And these are simply a list of priorities and a question of priorities.
So trade-offs, yes, but the value of those trade-offs clearly are dictated by how a stakeholder views their priorities and what they care about.
And I would just say that in the rather large nearly infinite universe of things that you can ask an LLM, you know, blocking off or potentially affecting some set of queries, you know, in general, will probably be a small price to pay.
Because, again, lots of people are anticipating different applications, and so, you know, there are -- for every single area where perhaps some subset of queries get affected, you could probably find many, many more where it has nothing to do with copyrighted content and, therefore, adding that safety measure will have absolutely no impact.
BY MS. MEHTA:
Q You mentioned the entire universe of kinds of things that LLMs can do, and I think you -- you

309

used a similar term about the universe of things.
What are you talking about?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I'm talking about the fact that users are unpredictable, and when you present a tool or an interactive, sort of, mechanism in front of them, it's difficult to anticipate what they're going to send it.
And so, you know, whether something is a good idea or not, whether it leads to good effects or good outcomes or not, users might -- you know, there's just very -- it's a difficult task to anticipate the entirety of all things that a user can or cannot ask, right?
So in some sense, we can do market research, we can look at user traces, we can do user studies that try to understand what users want out of certain systems and what they're likely to use it for, but, you know, characterizing that entirely is a -- is, sort of, a dynamic process.
And so what I'm trying to say is that anticipating and quantifying the exact cost of what potential queries might be affected and what

310

they -- you know, what the detrimental effects would be is difficult and also potentially small given the what seems like, you know, attempts by AI companies in general to grow the space of LLM queries.
Whereas, you know, on the other side of that trade-off, something that -- whether it respects copyright or not is something that is very concrete, that is very known, and the potential harms to creatives and those affected by copyright violations is also, generally speaking, quite concrete.
And so, yeah, I'm just pointing out that in a trade-off of sorts, it really speaks to, I think, the priorities at play when one side is a potential small piece of an unknown pie and the other is a very concrete known understanding of harms both legal and economic and personal.
BY MS. MEHTA:
Q You're not offering an opinion as to how the founders or executives at Anthropic have balanced those trade-offs in their own minds in terms of what they think their priorities are in various trade-offs relating to guardrails, are you?

311

MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A In general, no. I think I could be -- you know, it's late in the day, my head is a little bit fuzzy, but I believe there are some citations to prior testimony by Anthropic employees about their original goals and intents.
But, you know, aside from that, at a very general level, I would say that I don't know what their key priorities are except for cases where we have testimony that give us insight into it.
BY MS. MEHTA:
Q Would you agree that Claude has many uses other than, or outside of, uses relating to song lyrics or music?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I believe there are a number of uses that it's been marketed for. And I know that there are some uses where Anthropic has partnered with other companies, and so, you know, there are more specialized uses for it.
But, you know, I guess if you're asking do I understand the full extent of that? I do not. I'm not up to date on the marketing and the -- and

312

the, you know, business practices at Anthropic, so I can't speak with, you know, concrete certainty about the space of uses that they're targeting and trying to market.
BY MS. MEHTA:
Q Well, you looked at Clio data and talked about usage in your report. And, through that, you had an understanding that users use Anthropic's Claude models for a lot of things other than for song lyrics or music-related tasks, right?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A So, yes, I have looked at the Clio data. You know -- gosh, that's also been a while. Was that in the original report, I think?
You know, I don't recall concrete details of that other than that I discussed that at fairly length -- a fair bit of detail --
BY MS. MEHTA:
Q I can help you. I'm not asking for concrete details. Let me ask the question more simply.
You're not denying that users use Claude for many different things that don't relate to

345

You see that?
A Yes.
Q And then it says: [As read] "But Anthropic's inclusion of Publishers' lyrics in its training data also enables Claude to perform various other functions and use cases, including those relating to songs and lyrics specifically such as identifying songs based on lyrics, translating or modifying the lyrics to existing songs, generating, quote, new or, quote, original song lyrics in the style of a specific songwriter, artist, or song on a given topic, or otherwise in response to a user prompt or generating other output that includes parts of the original work blended with different texts."
Do you see that?
A I do.
Q Did you do anything in your reports to quantify the dollar value to Anthropic from any of those music-related tasks?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A So, just to clarify, you're asking me to tell you the internal Anthropic analyses of financial outcomes of their different applications

346

of their model?
BY MS. MEHTA:
Q No, I'm asking whether you included in your reports any expert opinion or analysis that would quantify what the value is that you believe that Anthropic has from copyrighted lyrics based on any of the tasks that you say that Claude can do based on Publishers' lyrics.
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A And, again, these kind of numbers would only be possible if I had access to Anthropic's internal documents about their costs, about their server costs, about their market opportunities, about the rates that they're charging individual users, the rates that they're charging for any potential licensees, the rates that are -- or, sorry, paying to any licensees, or any revenues that they're gaining from third-party application tools or companies that are using their tools.
So, no, I do not have that type of internal financial access to Anthropic, therefore, I cannot possibly make estimates of financial outcomes of any of these activities other than pure wild speculation.

347

BY MS. MEHTA:
Q And then as between these different activities, do you believe any one of them is more or less prevalent in terms of use of Anthropic's Claude?
MR. HOWELL: Objection. Form. Vague.
BY MS. MEHTA:
Q Based on data or evidence that you considered in this case.
MR. HOWELL: Calls for speculation.
BY THE WITNESS:
A I don't believe I speak to that in the report. If I do, you can point me to it, and we can discuss, but I don't believe I -- I speak to this.
MR. HOWELL: Counsel, we're at an hour and five minutes, I believe.
MS. MEHTA: Now is a good time for a break.
MR. HOWELL: Sure.
THE VIDEOGRAPHER: Going off the record at 3:59 p.m.
(WHEREUPON, a recess was had.)
THE VIDEOGRAPHER: This begins Media Unit 6. We are back on the record at 4:13 p.m.

348

BY MS. MEHTA:
Q Okay. Dr. Zhao, could you turn to Paragraph 71 of your rebuttal report this time.
MR. HOWELL: Should be Exhibit 2.
MS. MEHTA: Yes.
BY THE WITNESS:
A Sorry. What paragraph again?
BY MS. MEHTA:
Q 71.
So if you look at Paragraph 71, the first thing you say there is: [As read] "Third, Dr. Bilmes ignores what the, quote, use for Publishers' lyrics determined and informed" -- I'm sorry. I misread that. I'll start again.
"Third, Dr. Bilmes ignores that the, quote, use for Publishers' lyrics is determined and informed at least in part by what Claude users prompt Claude to do."
Do you see that?
A Yes.
Q Does Anthropic control what prompts Claude users input into Claude?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A No.

349

BY MS. MEHTA:
Q It goes on to say: [As read] "A Claude user that seeks copies of Publishers' lyrics is utilizing those lyrics for the same, quote, originally intended, end quote" -- sorry, "originally intended use, end quote, as its creators or copyright holders to enjoy the words to a song."
Do you see that?
A Yes.
Q Is it your view that the intended use of song lyrics is to enjoy the words to a song?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I think I'm sure there are many definitions, and different people have different takes. I think one of the purposes, yeah, is to enjoy the words to a song.
BY MS. MEHTA:
Q That's the purpose that you identified in your report, correct?
A Here, yes.
Q And is it your opinion that -- let's move -- abstract one level away from lyrics.
With songs generally, is the originally

350

intended use of a song to enjoy the song?
MR. HOWELL: Objection. Form. Vague.
BY THE WITNESS:
A I don't believe I've ever seen such a definition put on paper. Gosh, and, you know, to be more clear, even the definition of what a song is, you know, would obviously change the answer quite a bit. There are songs that don't have lyrics. There are songs that are primarily -- primarily lyrics-driven and melody-light and word-heavy.
And then -- yeah, so there's a whole variety of songs, and, gosh, everything from electronic music to ballads and country songs, which are, you know, of course quite heavy on storytelling and lyrics and a little bit light on, you know, background music and melody.
There's a whole range, and I suppose it really depends on the context and the person.
BY MS. MEHTA:
Q Right. I'm asking about all of those different types of songs, and I'm not separating out between the lyrics and the rest of the musical composition. I'm referring to the entirety of the musical composition.

351

You'd agree with me that, to use your words, the originally intended use of a musical composition is for people to enjoy the musical composition, correct?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I mean, as you questioned me about earlier this morning, I'm not a musicologist -- in fact, I'm not even sure what the definition of a musicologist is -- but I do believe there are other people who are more experts in this space who are participants in this case.
And I believe, although I haven't studied their declarations in detail, that they speak to issues like this --
BY MS. MEHTA:
Q Right.
A -- so I would defer to them.
Q But you've said here -- I mean, you did offer an opinion on this. You said: "A Claude user that seeks copies of Publishers' lyrics is utilizing those lyrics for same, quote, originally intended use as its creators or copyright holders to enjoy the words to a song."
Are you saying that you weren't qualified

352

to make that opinion?
MR. HOWELL: Objection. Form. Misstates prior testimony.
BY THE WITNESS:
A You're asking me a separate question about the intended uses of all songs. And when I specifically asked about and talked about the variety of songs and why they're functionally different and could be used in different ways, you insisted on a global definition, and that is the question that I'm deferring to a music expert on.
BY MS. MEHTA:
Q Well, let's go back to what you said here.
Do you believe that you are competent to offer an expert opinion with respect to what the originally intended use of lyrics to a song are?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I don't believe that I am representing that this is the only use -- only possible use for lyrics.
So, yeah, I believe this opinion is valid, and I stand by it that one of the purposes of lyrics is -- and one of its primary intended use purposes is to enjoy words to a song.

353

BY MS. MEHTA:
Q When you listen to music, is one of the reasons that you listen to music to enjoy the musical composition?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A The music composition, what -- what does that mean? Sorry.
BY MS. MEHTA:
Q I mean the whole -- the song. When you listen to a piece of music, is the intended purpose for listening to that piece of music to enjoy it?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I guess in general, yes.
BY MS. MEHTA:
Q If you could turn to Paragraph number -- you can go to your opening report.
A Okay.
Q And I'm going to have you turn to -- I'll get you there -- Page number -- you know, actually, let's first start with your rebuttal report at Paragraph Number 128. So this is Exhibit Number 2.

354

A Paragraph number again? I'm sorry.
Q 128.
In your -- and I'm going to ask you a general question first, and then I'll take you to Paragraph 128.
In your expert reports you opine extensively about what you believe to be the harms or risks associated with AI models, correct?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I'm -- I'm sorry. I wasn't quite ready, so I was shifting. Could you repeat the question?
BY MS. MEHTA:
Q In your expert reports, you opine extensively about what you believe to be the harms or risks associated with AI models, correct?
MR. HOWELL: Objection. Form.
BY THE WITNESS:
A I do.
BY MS. MEHTA:
Q Now, you understand that Anthropic has offered an expert to talk about some of the benefits of AI, Dr. Altman, right?
A Yes.
Q And in Paragraph 128, you say that: "The relevance of unrelated and disparate benefits of LLMs broadly or how Claude is being used outside of the context of this litigation is also unclear to me."
Do you see that?
A Yes.
Q Do you think it's reasonable, if the jury is going to consider the potential harms of AI that are unrelated or outside of the context of this litigation, do you think it's fair and reasonable for the jury to get to hear about the potential benefits of AI?
MR. HOWELL: Objection. Form. Vague. Calls for a legal conclusion.
BY THE WITNESS:
A I'm sorry. I was just thinking for a second. Could you repeat the question?
BY MS. MEHTA:
Q Yes. If the jury is going to consider the potential harms of AI that are outside of the context of this litigation -- so they have nothing to do with copyright regurgitation -- do you think that it is fair and reasonable for the jury to also get to hear about the potential benefits of AI?

356

MR. HOWELL: Objection. Form. Calls for a legal conclusion.
BY THE WITNESS:
A Is it fair? Gosh. I don't know, who is the arbitrator of fairness? I believe the judge here probably presumably?
I'm simply offering information on my opinions, and what is fair or not, I don't believe I give an opinion to that question. Yeah.
BY MS. MEHTA:
Q Well, in Paragraph 128, you say that: "The relevance of unrelated and disparate benefits of LLMs broadly or how Claude is being used outside of the context of this litigation is also unclear to me."
And my question is: Do you think that the analysis of benefits and harms of LLMs should be limited to the context of this litigation; or are you offering an opinion about harms of AI or LLMs outside of the specific issues in this litigation?
MR. HOWELL: Objection. Form. Calls for a legal conclusion.
BY THE WITNESS:
A I guess I'm not quite understanding the question.

397

essay that answers a specific question to the specifications that a single professor has given down for a particular class.

Now, today, that is trivially possible. That is no cost. A student can basically go literally for free to Anthropic Claude, and within presumably ten minutes, and quite possibly far less than that, they can produce and generate an essay detailed down to the exact specification of a particular homework assignment, take-home essay assignment, or whatever a particular professor has specified. And that act of plagiarism has been made so much easier by order, multiple orders of magnitude, given the presence and the free access to Claude and similar tools.

Now, if -- in the past we have of course dealt with plagiarism; however, all the mechanisms that we have to detect, defend, and prevent and mitigate issues of plagiarism, we have many.

They are mostly functional in the context of logistics where a student needs to expend significant resources to find a plagiarist -- a plagiarized essay. Whereas, now, it is trivially cheap and easy for them to do so, and they know that detection is potentially costly for us.

398

And, again, I come back to this question of this is about friction, this is about cost, this is about how easy it is and, therefore, how prevalent it is for certain bad actors or for people with bad intentions to carry out acts that produce harm.

MS. MEHTA: Move to strike. Not responsive.

BY MS. MEHTA:

Q When it comes to the benefits of AI, would you agree that the prevalence or scale of the benefits is also relevant?

You just went on a long soliloquy about how important the scale and the prevalence is when it comes to assessing the risks. It's also --

A No, I disagree.

Q -- it's also -- so you don't agree that we should also look at the prevalence or scale when it comes to the benefits?

MR. HOWELL: Objection --

BY MS. MEHTA:

Q Is that -- is that your testimony?

MR. HOWELL: Sorry --

BY THE WITNESS:

A Are you done with your question --

399

MR. HOWELL: Let's --

(Unreportable cross-talk.)

MS. MEHTA: Let me start the question again.

MR. HOWELL: Go ahead.

BY MS. MEHTA:

Q When it comes to the benefits of AI, it's also important to look at the prevalence and the scale of those benefits in evaluating the cost/benefit analysis; isn't that right?

MR. HOWELL: Objection. Form. Vague.

BY THE WITNESS:

A I believe that looking at benefits of AI requires, just as it does to look at harms, to look at concrete examples, quantifiable data, and real research that studies the benefits and harms of AI.

BY MS. MEHTA:

Q And --

A I'm not -- sorry. I'm not done --

Q Go ahead.

A I'm not saying that AI is all bad and has no benefits; there are benefits to AI. There are benefits to multiple types of AI.

And even for a more constrained definition

400

of generative AI, I do believe that there are scenarios where they have some benefit. However, understanding of those benefits should be done with analysis, research, and actual quantifiable data rather than single anecdotes of: I saw one instance of this benefit and, therefore, it must generalize to all.

When I speak to -- about harms, I'm referencing harms that are known, that are not denied, that are observed by many, that are reported in public press, that are reported in academic journals, that are reported in peer-reviewed research publications from many different sites. And there's little dispute that these reports are real, that these harms are together today present today, not hypothetical, not speculative.

BY MS. MEHTA:

Q Have you cited in your reports any public press, academic journal, or peer-reviewed research publication about the benefit of AI?

MR. HOWELL: Objection. Form.

BY THE WITNESS:

A I will be frank: I have kept an eye out for such articles. I have looked for concrete

401

evidence of applications where AI has provided benefit. I have not been able to see concrete evidence in terms of peer-reviewed papers that are convincing to me.

I acknowledge that there are certain applications where, even by speculation or reasoning about AI's functionality, that it would prove to be positive and useful.

Some of these domains include, for example, you know, interior decorating, right? For example, a tool that would interact with you and make suggestions about how your house might be laid out, about what would be an attractive, you know, arrangement of your interior of walls, of items and things of that sort. And that would absolutely be a positive.

And, in particular, you know, it's important to note that there are applications where the harms or the potential mistakes made by AI, whether it be memorization or regurgitation, there are some domains where those things are less meaningful or impactful.

So interior decoration is one of those examples. If I don't like the layout that's given to me by an LLM because it made a mistake and the

402

couch is halfway through the wall and sticking out halfway, that's okay because something like that is not harmful to me other than a small annoyance and I can hit reload, and presumably it will try again and maybe get it correct.

There are other scenarios where I believe, talking from some of my colleagues and having -- not necessarily reading their publications, but having conversations where I can believe that there are other benefits. Perhaps, for example, you know, in some applications of data searching, you know, in some instances where -- where we are basically searching blindly for data, you know, if an LLM or another similar tool can give some preference about maybe you should look in this direction first, that, again, is an example that, you know, to me I'm unsure about how much benefit there is.

However, I can convince myself that the harms are minimal in the sense that if we're in a domain where it's already random, perhaps trying in one random sector versus another, the harm is not significant because, at worst, you are just doing a slightly version -- slightly different version of random.

403

So it's a, perhaps, waste of resources, but, you know, the stakes there are not high and there's no obvious harm to humans or the scientific process.

BY MS. MEHTA:

Q This is a yes-or-no question: Before you submitted your expert reports in this case, did you do a literature review to look for peer-reviewed research pieces in the academic journals or public press that talked about real concrete examples of the benefits of AI; yes or no?

MR. HOWELL: Objection --

BY THE WITNESS:

A Yes.

MR. HOWELL: -- to form and to the instruction.

BY MS. MEHTA:

Q And your testimony is that you could not find any peer-reviewed research, academic journals, or public press reporting that talked about the benefits of AI? That's your testimony?

MR. HOWELL: Objection. Form. Misstates prior testimony.

404

BY THE WITNESS:

A My testimony is that I did look for popular press and news and articles about benefits of AI, and the articles that I found, nearly every time when I looked at it and I dug a bit further, was less than convincing.

I will also say that I did not spend a ton of time looking for negative harms. This was not a -- you know, I didn't spend, you know, a tremendous amount of effort digging around looking for -- for case scenarios of bad things that may or may not be true.

The harms that I discuss, in large part, come from or stem or I was clued in into these issues by first -- first-hand experience either in my workplace with the educational impact or by conversations I've had with experts in the field.

In some cases it's been spurred on by conversations with people who are directly impacted by the harms. In many cases, whether it's artists or writers, I can give you at exhaustive length -- actually, I can't give you full exhaustive length, but it's quite a long length -- list of people who have been harmed directly by these models who I have had firsthand

461

CERTIFICATE OF REPORTER

I, MICHELLE M. YOHLER, a Certified Shorthand Reporter within and for the County of Will, State of Illinois, do hereby certify:

That previous to the commencement of the examination of the witness, the witness was duly sworn to testify the whole truth concerning the matters herein;

That the foregoing deposition transcript was reported stenographically by me, was thereafter reduced to typewriting under my personal direction and constitutes a true record of the testimony given and the proceedings had;

That the said deposition was taken before me at the time and place specified;

That I am not a relative or employee or attorney or counsel, nor a relative or employee of such attorney or counsel for any of the parties hereto, nor interested directly or indirectly in the outcome of this action.

462

IN WITNESS WHEREOF, I do hereunto set my hand and affix my seal of office at Chicago, Illinois, this 24th day of February, 2026.

Michelle M. Yohler

Michelle M. Yohler, CSR, RMR, CRR
Certified Shorthand Reporter
CSR No.: 84-4531

No. 621063

Re: Deposition of **Ben Zhao, Ph.D.**
Date: 2/20/2026
Case: Concord Music Group -v- Anthropic
Return to: transcripts@planetdepos.com

Errata

Page: Line	From	To
10:4-6	It's a long claim with a lot of complaints that I read three years ago.	It's a long **complaint** with a lot of **claims** that I read three years ago
26:19-21	I'm not going to get into my personal opinion about what are these issues called.	I'm not going to get into my personal opinion about what **all** these issues **are**.
36:18-19	I meant that if I have a form – fully formed definition	I meant that if I **had a fully** formed definition
81:79-81	we make sure that there is no offensive or legal content	we make sure that there is no offensive or **illegal** content
132:24	what current models do, and current models behave	what current models do, and **how** current models behave
177:15-17	The definition of a data retrieval system is incredibly expensive	The definition of a data retrieval system is incredibly **expansive**
188:23-25	There's cureration, there is weighting, there is all sorts of things that may or may not happen	There's **curation**, there is weighting, there is all sorts of things that may or may not happen
190:20	cureration	**curation**
191:1	tokenize table	**tokenizer** table
194:18	cureration	**curation**
222:25-223:3	I'm sorry. I assume including that is with the research documents but it also includes things like public press and, you know, newspaper and whatnot.	I'm sorry. I assume **included in** that **is the** research documents but it also includes things like public press and, you know, newspaper and whatnot.
289:1:	cureration	**curation**
293:17	cureration	**curation**
293:23	cureration	**curation**
294:3	cureration	**curation**
295:7	cureration	**curation**
295:10	cureration	**curation**
297:12	cureration	**curation**
297:15	cureration	**curation**

No. 621063

Re: Deposition of **Ben Zhao, Ph.D.**
Date: 2/20/2026
Case: Concord Music Group -v- Anthropic
Return to: transcripts@planetdepos.com

Errata Cont'd

Page: Line	From	To
328:6	cureration	**curation**
449:19	cureration	**curation**
452:7	cureration	**curation**

4/1/26
(Date)

(Signature)

No. 621063

Re: Deposition of **Ben Zhao, Ph.D.**
Date: 2/20/2026
Case: Concord Music Group -v- Anthropic
Return to: transcripts@planetdepos.com

ACKNOWLEDGMENT OF DEPONENT

I, Ben Zhao, 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.

4/1/26 _____
(Date) (Signature)