EXHIBIT 139



HIGHLY CONFIDENTIAL--ATTORNEYS' EYES ONLY

Transcript of Andrew Bray

Date: February 24, 2026
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
NORTHERN DISTRICT OF CALIFORNIA
SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL., Plaintiffs, v. ANTHROPIC PBC, Defendant. Case Number 5:24-cv-03811-EKL SVK

HIGHLY CONFIDENTIAL--ATTORNEYS' EYES ONLY
VIDEOCONFERENCED and VIDEO-RECORDED
DEPOSITION OF
ANDREW P. BRAY, Ph.D.
San Francisco, California 94111
Tuesday, February 24, 2026

Reported Stenographically by:
MARY J. GOFF
CSR No. 13427
WA CSR No. 21030779
Job No. 620584
PAGES 1-225

2

HIGHLY CONFIDENTIAL--ATTORNEYS' EYES ONLY
VIDEOCONFERENCED and VIDEO-RECORDED DEPOSITION of ANDREW P. BRAY, Ph.D., Volume I, taken on behalf of Plaintiff Publishers, Oppenheim + Zebrak, LLP, at WilmerHale, 50 California Street, Suite 3600, San Francisco, California 94111, beginning at 9:03 a.m. and ending at 4:10 p.m., on Tuesday, February 24, 2026, before MARY J. GOFF, California Certified Shorthand Reporter No. 13427 and WA CSR No. 21030779.

3

APPEARANCES:
For Plaintiff Publishers
Oppenheim + Zebrak, LLP
BY: KEITH HOWELL
Attorneys at Law
4530 Wisconsin Avenue
5th Floor
Washington, D.C. 20016
202-957-5043
khowell@oandzlaw.com

For Defendant Anthropic PBC
WilmerHale
BY: DISHA PATEL
LOUIS W. TOMPROS
Attorney at Law
60 State Street
Boston, Massachusetts 02109
disha.patel@wilmerhale.com
617-526-6795

ALSO PRESENT: Natalie Naugle, in-house counsel Anthropic
Videographer: Justin Ramirez-Hale

4

INDEX

WITNESS	EXAMINATION
ANDREW P. BRAY, Ph.D.	
Volume I	
ATTORNEY HOWELL	6
ATTORNEY PATEL	217

NUMBER	DESCRIPTION	PAGE
Exhibit 1	Anthropic's Supplemental Responses to the Publisher Plaintiffs' First Set of Interrogatories, No. 4	36
Exhibit 2	Lyrics ANTH_SAMPLE000048259	139
Exhibit 3	Lyrics Viva La Vida ANTH_SAMPLE000217578	175
Exhibit 4	Exhibit B to Publishers' Amended Complaint	179

Pages 49-52

49

the regular expression matched to, but that clearly do not -- were not in this case seeking any amount or form of lyric regurgitation.

Q And so you -- you recognize that there are false positives in the list of files that match your regular expression?

A That's correct.

Q You didn't remove those files from the list of documents that you found to be responsive to each research question correct?

A That's correct. They remain in the -- in the collection of documents that were responsive to that regular expression.

Q Why wouldn't you remove false positives?

A The approach that was used here is a standard approach in building a classifier model.

Whenever you're building a classification model -- a two-class classification model, there's the possibility that that classifier will generate false positives. There's the possibility it'll generate false negatives.

It's, in many cases and certainly in this case, impossible to write a classifier -- realistically write a classifier that could do perfect sorting that wouldn't have either false

50

positives or wouldn't have false negatives, so the question then becomes: How flexible should my model be -- work -- can I build a model? And here by "model," I mean the regular expression that will capture the -- the sort -- sort of text documents that I'm seeking without incurring too many false positives.

And so it's trying to balance the -- as proportions, these are the recall and the precision of the classifier, in this case thinking about just having a balance of false positives and false negatives.

Q Do you have a sense for how many false positives there are in your set of documents for Research Question 1?

A I haven't reviewed enough of them to have kind of a systematic estimate. I would -- I think it's fair to say just many of the documents are false positives.

Q Can you provide any more precision regarding what percentage of the records you found to be responsive to Research Question 1 you considered to be false positives?

A I wouldn't provide any more precision. It was based on a spot-check and the determination that

51

indeed there were still false positives in the -- you know, upon seeing false positives, there's the option of making the regular expression less flexible.

So the original regular expression was not very flexible. It had only matched to a few very explicit types of files. And the goal here was to have a classifier that had high precision, which is to say of those that we were classifying as positive, we wanted to be sure that a high proportion of them were positive.

The second stage -- and this is what describes the refinement of the regular expressions -- this was to address the recall side of the equation. So initially the model was built to be kind of deliberately strict, knowing that it would be excluding certain ways that people would, you know, write down a -- when they're seeking any amount or form of lyric regurgitation, knowing that we are -- you know, there are more subtle forms of asking for -- more indirect forms of asking for it.

And knowing that we would miss those in the original regular expression, I had Claude -- we can get into this more explicitly, if you would like -- surface examples of these false negatives

52

that allowed me to find ways that I could broaden that regular expression.

So it's a -- you have these competing forces. You have -- if you have a lot of false positives, there's a -- kind of a need to constrain your regular expression, make it a little bit more strict. If you have it too strict, you are going to be having more false negatives, and so there was a back-and-forth trying to determine the -- the -- kind of the Goldilocks zone here of regular expression that was just the right level of flexibility that would capture most of the text documents that were seeking any amount or form of lyric regurgitation without inducing a flood of false positives.

Q Do you have a sense for the false positives rate for Research Question Number 2?

A So Research Question 2 is looking at those that were responsive to Research Question 1 and then further refining kind of what it was looking for. It's looking for the, I believe it's, song title or artist.

And if -- if it found those within particular parameters of the regular expression, it was responsive. In terms of the -- the -- similarly

Pages 61-64

61

A Right. Thank you.
So yeah, first being clear what a false negative was, there's an expression that is not currently being captured by the regular expression. That notion informed the original formulation of the regular expression. So that -- that first form that shows up in -- it's not labeled View 1, but under the Q1 section, that more paired back regular expression, you know, even that is keeping in -- it has some amount of flexibility built in to catch as many genuine expressions of this behavior that we're seeking to capture.
But beyond that, the main tool that was used to address the possibility for false negatives and wanting to just the regular expression accordingly was the -- this process of Claude refinement.
Q Sorry. You said the Claude refinement was the main tool that you used to assess false negatives. Was -- anything else in particular you used to assess false negatives associated with your regular expression?
A I think, you know, limiting the false -- addressing false negatives, wanting to capture -- you know, multiple ways to express this particular

62

behavior, you know, that -- and that -- that was the goal of part of the regular expression. The other part being restricting the false positives.
And the way that that concern was materialized was in, you know, my own formulation of that first version of the regular expression and the subsequent refinement using Claude.
Q Okay. For Research Question 1, are you able to quantify how many false negatives there are in the sample that you didn't find?
A No. So the extent to which, you know, false negatives -- the way that I dealt with false negatives was through kind of holistic, not data informed -- or not directly data informed, not algorithmically informed, articulation of the regular expressions in the process of Claude refinement.
But the -- you know, an assessment of the false negative rate would require -- in this case because the regular expression were run over, I guess, roughly 2.5 million, the adjusted sample, it would require an exhaustive -- like, actually knowing that rate would review an exhaustive review of those records. And doing, you have to consider -- one -- I guess how many -- which records

63

do you look at? Do you look at all of them? Do you look at a subset of them in trying to assess that error rate?
And then secondly you look: How do I determine what the ground truth is?
So you know, the notion of a false negative is premised on the notion that, like, I can -- I can show you a file and I can tell you beyond a shadow of a doubt that that expresses -- that should have been responsive, because that does capture this thing that I'm trying to capture.
And you know, methods that you could try to get at that ground truth are, I would say, manual review. That's one method.
Another method is kind of an analog using large language models to do that review. My experience working with Claude in that process is that, you know, it -- it was not providing me the ground truth. It was providing suggestions.
I disagreed with its suggestions sometimes. I -- it flagged certain -- certain files that were not kind of concordant with what I was seeking in those research questions.
Similarly, the manual method that was used by Mr. Dennis, that, too, is shown to be

64

error-prone. And you know, he has -- he does his own error analysis. And then Mr. Lustig does an error analysis on -- on that error analysis and found that even that manual process had its own errors.
So the difficulty of establishing what a ground truth would be, particularly at scale in order to estimate that false negative rate, would give kind of a misleading sense of -- of certainty because it's -- it's -- as the -- this case has shown and Mr. Dennis's categorization followed by Dr. Lustig's, it's -- it's very difficult to determine for any particular file whether or not it should be responsive to a particular research question.
Q So it's practically impossible to determine the false negative rate for at least Research Question 1?
ATTORNEY PATEL: Objection, mischaracterizes the witness's testimony.
A I would say the -- so quantifying the false negative rate would require a sense of ground truth. And practically getting a ground truth -- so using the method that Mr. Dennis used, which is to say manual review followed by -- well, I will move

Pages 141-148

141

if it was -- if it fell into these two classes of documents that in the report I call "false positive or false negative disagreement between the original regular expression and Claude," then yes, I would have reviewed this.

Q Okay.

A And it looks familiar, because it came up again in Mr. Dennis's report.

Q Do you have an opinion one way or another if this -- if the prompt here, "sing me a Lady Gaga song," if that is a request for lyrics?

ATTORNEY PATEL: Objection, form.

A So what I was seeking to determine here was the -- whether or not the prompt was seeking any amount or form of lyric regurgitation. And in this case as I read this, "sing me a Lady Gaga song," I don't see a request for lyrics. I see a request for singing a song.

Q (BY ATTORNEY HOWELL) Do -- you see that Claude provided lyrics to a Lady Gaga song in response to that -- this prompt, correct?

A Right. I can see the response that Claude provided, but the intent of Question 1 and Question 2 didn't have to do with Claude's output, but had to do with the user's input; and so there

142

the determination is just based on the phrase "sing me a Lady Gaga song."

And to me, that's expressing -- well, exactly what it says, that -- asking the -- Claude to sing a song. And I can imagine lots of possible responses that Claude might give that wouldn't involve lyrics -- the provision of lyrics as we see it here.

Q But you understand that Claude is a text chatbot?

A I understand that Claude is a text bot -- a --

Q -- so --

A -- chatbot, yeah.

Q -- any response from Claude would -- singing a song would have to be in the form of text?

A It's not --

ATTORNEY PATEL: Objection, form.

A -- it's not clear to me that that's true. So Claude could, for example, provide a link to -- it could -- it could respond in text, saying: I'm an LLM that provides texts and not songs, and therefore, I cannot sing you a song.

I can imagine it could provide text that says maybe that same thing, but: Here I can link

143

you to a video of a 14-year-old of YouTube singing a Lady Gaga song.

Both examples of, you know, text output that aren't lyric regurgitation on the output side as I was -- would characterize it. But again, this was -- the determination had to be made based on the intent of the user, not on the response of the LLM.

Q And your determination was this user did not seek any amount or form of lyric regurgitation?

A That's correct. It didn't fall into the definition of "lyric regurgitation" here; that they were asking the LLM to sing a song.

Q So this would not have been a document that you looked at and then attempted to modify with your regular expression in order to cover cases like this?

A That's correct. You might call this a "false" false negative, which is confusing.

Q And if you -- there's various headers here. Do you see "prompt, human, assistant, completion"?

A Um-hum.

Q Do you see those?

"What follows prompt," that's the system prompt, correct?

144

A That's my understanding; that it -- at the top of each one of these files is the system prompts that precedes the actual prompt that a user supplied.

Q And what is your understanding about what a system prompt is?

A I guess more or less just what I said; that it's a prompt provided to a large language model that precedes a user's text or specific query. And it's taken into the context of the model into the context window.

So for example, this system prompt allows a human to ask a question like, "What is the weather today," and it -- or "what happened three days ago in New Orleans?"

And the -- because the model actually has access to the date through the system prompt, you would have some means to yield a useful answer.

Q Okay. And the date that's included in the system prompt, that's the date the last prompt in a given record was submitted to Claude?

A That's my understanding.

Q Okay. And you said Claude relies on that information to provide answers that might be dependent upon the date?

145

A Dependent on the date, depending on -- on other -- all the information. You know, the context is taken into account. So the -- just by starting this text document off by saying "the following is a conversation between a human an OpenAI assistant," the LLM will be inclined to make that actually become true by having an alternating dialogue between humans and AI.

Q And you also reviewed metadata associated with records in the sample, correct?

A That's correct.

Q And one of those metadata fields is a field called "Prompt Created Date CF"?

A Right.

Q Do you recall that?

What's your understanding of what field conveys?

A My understanding of "Prompt Created Date CF" is that it's the -- the closest timestamp that we have to when the last prompt was submitted to Claude. It's the closest timestamp that we have to the actual creation of that file.

Q And you said the "System Prompt Date" is the date on which the prompt was -- the last prompt was submitted to Claude?

146

A So -- yes, but here, I guess, it's useful to be a bit more precise. The "Prompt Created Date CF," that field is measured in UTC, you know, a standard -- a time standard that works regardless of whatever time zone somebody is in.

At as you can imagine, if the "System Prompt Date" followed that same rule, then somebody asking -- what is it -- late at night here in California for a prompt that used "Prompt Created Date CF," it might be the next day in London, and so the date actually would not be the date at which the local user was sub -- providing the prompt.

And so because of that, these prompts, I understand -- sorry. The dates that are found in the system prompt are not the date -- the time UTC of the last prompt, but the time in -- the local time of the user at which that last prompt was submitted.

Q Okay. That's helpful.

So the difference between the "System Prompt Date" and the "Prompt Created Date CF" is the difference between -- the time difference between where that user is located and UTC, the time zone difference, essentially?

A More or less. As long as you're restricting -- your interpretation of date is the date of the last prompt. Any preceding prompts in that text sequence, you know, is bounded by that date, but could well have been beforehand.

And I came across examples of files that appeared to be from the same conversation, which is to say you see the first few exchanges of a conversation in one file, the next few con -- snippets of conversation in the next that subsume the previous one, and so on.

And that if you look at the "Prompt Created Date CF" or the "System Date" of that smallest file earliest on the conversation, that could lag -- could precede the last date by at least months in the example that I have seen.

So that's just to say the interpretation -- both have to keep in mind time zones when understanding the date in which the files were created, and you also have to understand the nesting structure.

And neither of these seem to have been taken into account by Mr. Dennis in his report, which is why the -- any part of his analysis that bears on kind of a temporal analysis has significant flaws in it.

148

Q Given that understanding, the Prompt Created Date -- sorry --

The Prompt Created CF Date and the System Prompt Date for a given record should never be more than one day apart?

A I guess I would have to think through the full -- in general, I would say that -- that's roughly true.

Like, you know, there are daylight savings time, and you know, that can sometimes add or sub -- but it's on the order of a day or maybe a little bit more depending on if there are certain files that happen to span one of these calendar switching moments.

Q Never more than two days?

A That's the -- the thing that I would say would never be more than two days would be the difference between the date found in the System Prompt and that Prompt Created Date CF of the same file, both of which referred to the time in which the last prompt was submitted.

Q And if you go to paragraph 139 of your opening report.

A Great. Um-hum.

Q Your opening report discussed various

Pages 217-220

217

witness.
ATTORNEY PATEL: So before I begin questioning, I just wanted to put on record that we're designating this transcript as "Highly Confidential" pursuant to the protective order, and that we reserve the right for the witness to read and sign the transcript pursuant to Rule 30(e)(1).
EXAMINATION
BY ATTORNEY PATEL:
Q Dr. Bray, you were asked whether you performed any analysis to confirm that the record universe contains at least one record for every prompt submitted to Claude.AI or Claude's API.
Do you remember those questions?
ATTORNEY HOWELL: Objection, form.
A So here you're referring to a question from Mr. Howell?
Q (BY ATTORNEY PATEL) Yes.
A Yes.
Q Could you please turn to page 4 of your rebuttal report? On page 4 of your rebuttal report, you include heading that says "Ignoring the Nested Structure of Text Data," right?
A Yes, I see that.
Q What does the "nested structure of text

218

data" mean?
A This refers to the manner in which data is sent from a user to an LLM and then the manner in which an LLM provider stores the data that at the beginning of a -- what you might call a conversation, a multiprompt and output back-and-forth.
The first request that's sent from the user to the LLM actually contains two things. It contains a system prompt that sets a baseline of the conversation that we have seen in these files. It also contains the original user prompt.
On Anthropic's end in this case, the language model will produce an output. And then that collection of the system prompt -- the prompt in the output is put into a single file.
Now, the second part of the conversation involves the user sending a follow-up question through their software to the LLM. That request now, it's not just containing their second prompt. It contains the original system prompt, their first question, Claude's response, and then their second question.
On Anthropic's end you'll have the large language model produce an output, and then a data

219

file will be created that logs now in this case the system prompt, a prompt and output, a prompt and an output.
And so by "nested structure," I mean that you have a conversation that in this case lasts two rounds, prompt output, prompt output.
And it corresponds to two files, but those two files have an overlap. And the later file is a strict superset of the earlier file. It contains all the contents of the first, plus the additional content of the second interaction.
And this goes on and on and on so that a conversation that lasts a hundred back-and-forths will have a hundred corresponding data files. And actually, all 100 of those would have original prompt visible.
Q (BY ATTORNEY PATEL) Have you seen examples of prompts and output pairs -- prompt and output pairs from a single conversation appearing more than once in the five million sample that you reviewed?
A I have.
ATTORNEY HOWELL: Objection, form.
Q (BY ATTORNEY PATEL) Why would prompt and output pairs from a single conversation appear more than once in the five million sample?

220

ATTORNEY HOWELL: Objection, form.
A So there are a few reasons here. I can think of three right now. The first relates to nesting. So exactly what I just described.
As a -- as a conversation precedes, the initial components of the conversation are propagated into every subsequent file. So those are duplicated. Those components are duplicated. And that's one source of duplication.
Another source of duplication is that -- I believe this is after the user sends their first prompt. Two things will happen, at least.
One is that Claude will provide the output. The other is that a model will be triggered -- or a process will be triggered that sends an independent query to Claude, asking Claude to provide a title that summarizes kind of the -- that first back-and-forth.
And that would create a title record. So that initial prompt, actually, in that case would exist in the title record, as well as in the record that records the prompt and the output.
There's a third source of duplication. And that is the -- this process of duplicate LogIDs where, to my understanding, a LogID should be unique