EXHIBIT 2

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

5

San Francisco, California 94111
Tuesday, February 24, 2026
9:03 a.m.

THE VIDEOGRAPHER: Here begins Media Number 1 in the videotaped deposition of Andrew Bray, in the matter of Concord Music Group, Incorporated, 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 24, 2026. The time on the video monitor is 9:03 a.m. or -- yes. The videographer today is Justin Ramirez-Hale, representing Planet Depos.

This video deposition taking place at 50 California Street, Suite 3600, San Francisco, California.

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

ATTORNEY HOWELL: Keith Howell from Oppenheim & Zebrak on behalf of the Music Publisher Plaintiffs.

ATTORNEY PATEL: Disha Patel from Wilmer Cutler Pickering Hale and Dorr, counsel for Anthropic. And with me we have Louis Tompros, also

6

from WilmerHale; and Natalie Naugle, counsel at Anthropic.

THE VIDEOGRAPHER: The court reporter today is Mary Goff, representing Planet Depos. The witness will now be sworn.

THE COURT STENOGRAPHER: My name is Mary Goff. I am a California Certified Shorthand Reporter, CSR Number 13427. Today's proceedings are being captured by stenographic means.

ANDREW P. BRAY, Ph.D.,
Being first duly sworn or affirmed to testify to the truth, the whole truth, and nothing but the truth, was examined and testified as follows:

EXAMINATION

BY ATTORNEY HOWELL:

Q Good morning.

A Good morning.

Q Could you please state your name for the record?

A Andrew Bray.

Q And do you prefer Dr. Bray? Professor Bray?

A Either is fine. Probably Dr.

Q And you understand you're here today to have your deposition taken, correct?

7

A I do.

Q And it's -- it relates to the case "Concord v. Anthropic"?

A Correct.

Q Okay. And you submitted four reports in this case, correct?

A That's correct.

Q And you were retained by counsel for Anthropic?

A I was -- right. I was retained by WilmerHale to serve -- provide an opinion in this case.

Q And who drafted your four reports?

A I drafted the four reports -- let me back up.

The reports were made in response to questions provided to me by WilmerHale of areas that they wanted me to look into and render opinions on. This had to do with lyric regurgitation.

I devised a method of analysis and conducted the analysis and drafted the reports. And some of the drafting I directed counsel to assist with, and they assisted with that drafting.

And then I would review their drafts and revise anything that I disagreed with. And then I

8

reviewed the extent of the reports before signing them.

Q Okay. Other than counsel for Anthropic, did anyone else assist you in drafting your reports?

A So in the drafting of the reports, it was -- there were no others involved. But indirectly, there were others involved that informed the reports. There were other expert witnesses in the case that I spoke to on occasion.

And for about one month I worked with a research associate who helped with the computing side of things. This was from about -- this was during January.

Q And did that research associate contribute any language to any of your reports?

A He did not.

Q And what's their name?

A Calvin Carter.

Q And which portions of your reports did counsel draft?

ATTORNEY PATEL: Objection, mischaracterizes the witness's testimony.

Q (BY ATTORNEY HOWELL) You can go ahead.

A So the drafting of the reports was done jointly in terms of the original draft, the

13

Q Have you ever worked for a generative AI company before?
A I have not worked for a generative AI company.
Q And would that include both direct employment or -- and any consulting work?
A Certainly not direct employment. There's a company who I have done some consulting work for who are, since I have done consulting work for them, starting to get into areas related to generative AI.
Q But the areas you worked on with that company didn't address generative AI; is that fair?
A Not directly. This is a -- it's a software company dealing with data and models, but not LLMs in particular.
Q Given it was only one case, I think the answer probably "No," but have you ever been retained by counsel for Anthropic in any prior cases?
A I have not.
Q Okay. Have you ever published any research on generative AI?
A So the CV, as it was provided, in that CV there's no publications about generative AI. There on other areas of statistics and -- and

14

classification models, models for data. However, there's a paper that is -- a research study that I'm currently engaged in that does bear on that, but the manuscripts and draft phase has not yet been published.
Q Okay. So that manuscript would not be publicly available; is that fair?
A That's fair.
Q Okay.
A We have -- I should say we have submitted -- there are possibly public grant proposals that describe the -- the research.
Q Okay. If I wanted to look for one of those grant proposals, what search terms would I use?
A "Grading accuracy" would be the main one. So the research, it's looking at -- one of applications that's come up for large language models is in grading student work. This is something that's generally done by humans where you have an expert, a teacher devising a rubric for a particular question, and then you have a human -- at Berkeley we use students to apply rubric to student work.
And this has been kind of the only way

15

that we can get grading done at scale. And it has its benefits. It has its downsides. We have perhaps an alternative through LLMs that perhaps LLMs can be useful in this task of grading student work. And so we're simply looking at the question of accuracy of -- the accuracy of student graders versus the accuracy of LLMs as -- as graders. How do they compare, and in what cases do LLMs outperform and in which cases do humans outperform.
Q And do you have any published research where you use regular expression?
A I would have to -- can I take moment --
Q Sure.
A -- to look through? So I think my CV was the first. I can't state for certain, but it's possible that the 2019 paper, "Nature Climate Change," involves regular expressions.
Q Which page of your CV are you looking at?
A This is the first page.
Q This is under "Refereed Articles"?
A Correct.
Q And I'm sorry. What was the --
A The title is "Thermal Extremes Drive Fish Die-offs in Temperate Lakes."
I'm uncertain. The flag for a regular

16

expression would be kind of the -- if -- it was in an asserted process text data. And the data that we were looking at there was coming out of fishing game out of the Upper Midwest, and it may have been necessary to apply some text processing in order to analyze that data.
Actually, there is -- there's a piece of software here that I believe uses regular expressions. And that is under -- on the next page, page 2, "R packages, Jayaprakash, Voutov, and Bray, 2024, nemogb."
Q Okay.
A So again, that's a -- this is piece of open-source software that processes text data, and so regular expressions are -- were used to do that in a structured manner.
Q Okay. And for the article, I think you said that article could have used regular expressions, but you weren't sure; is that --
A I wasn't sure, yeah. And it would be in the data processing step, so it almost certainly wouldn't appear in the text of the article itself. It would have been a component that enabled the analysis.
Q All right. Any other articles involving

17

use of regular expression?
A Yeah, it's possible -- I mean, I'm -- it's possible that the previous one, the -- or the subsequent one, I guess, "Zhao, 2021 Global Multimodel Projections of Local Urban Climates." I'm looking for papers that apply the analyses that -- where the data could have had text fields. And in those circumstances, regular expressions likely would have been the approach that I took.
Q Okay. Anything else?
A I believe that's it.
Q And have you ever testified before?
A I have. I have testified in the -- in a case regarding ballistics.
Q And that was a criminal case, right?
A I believe so, yes.
Q Were you deposed?
A I -- I don't recall being deposed.
Q You testified at a hearing?
A I testified at a hearing.
Q And that was just on one occasion?
A That is correct.
Q And when were you retained in this matter?
A I believe it -- I was retained by WilmerHale in September.

18

Q September of?
A Of 2025.
Q Okay. And how many hours did you spend preparing for today's deposition -- specifically preparing for today's deposition?
A Probably on the order of 20 hours. I can't be quite sure.
Q And how much time have you spent in total on this case since being retained in September of last year?
A I'm not sure. I don't have that number in front of me. If I were to estimate, I would say for the opening -- everything that led up to the opening report. So between September and December 19, which I believe was when it was served, on the order of 70, 75 hours, and then with most of that coming towards the end of the time period. For the rebuttal report, maybe 50 hours. And then the reply report, maybe not 50 hours. Maybe a bit less for the reply report. Maybe 30, 40 hours.
Q Okay.
A It's all just ballpark figures.
Q And then the supplemental report, do you recall?

19

A That was brief. Maybe six hours.
Q And you mentioned your rebuttal report. You -- your rebuttal response to an analysis performed by Mr. Dennis, correct?
A That's correct.
Q Did you review any other reports submitted by experts retained by the Music Publishers outside of Mr. Dennis, I should say?
A Yeah, I guess as I was -- I'm unsure. The primary report that I was consulting with -- certainly Mr. Dennis's. The reason why I'm unsure is I reviewed lots of documents to get the context. These are documents that Mr. Dennis cited. And I guess I -- it wasn't -- it's not clear to me whether the author of those reports, who they were retained by or precisely what role they played in this case.
Q Okay. Do you know if you reviewed any expert reports in -- submitted in this case by experts retained by Anthropic?
A I can recall two reports, I believe, from experts retained by Anthropic. One was Julie Davis, providing a -- I believe it was her supplement. And this was -- provided information about the date at which the users in a particular set of users had become subscribers -- paid subscribers. So I reviewed that, and that informed the analysis that was presented in my supplement. I also reviewed a portion of Mr. Lustig's analysis. I'm not quite sure which report exactly this was, but it gets to the -- the question of -- it's cited in Dennis's report, but the reliability of his categorizations.

20

Q Okay. And so did you form any opinions based on your review of Dr. Lustig's report?
A I don't believe that my review of Dr. Lustig's report I don't think any of the -- hum. Can I check it? I'm trying to recall if it made it into these reports or if I would have -- I don't recall when I saw it. So if you would give me a moment. It would, I guess, probably be in my reply. Actually, let me see the rebuttal. That's the rebuttal. So I don't believe that either -- I don't believe that Mr. Lustig's report informed any of the opinions expressed in the reports that I have submitted. But Julie Dennis's -- Davis. Julie Davis's report informs my -- the opinion that was expressed in my supplemental report.

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

65

to say manual review, that process would be challenging.

But the even -- even after the -- like, the practical aspects of it were dealt with, which is to say just the volume of the data, you run into the problem of: We don't have a reliable ground truth model here. Right. We have -- we have, you know, multiple opinions by multiple humans, Mr. Dennis, Mr. Lustig, discussing the questionable state of just two documents.

And in this case the -- to get that false negative rate, you would have to -- you should have that level of scrutiny over 2.5 million documents. And even then there's still some debate over: For every given file, is that actually -- what is it capturing?

So the advantage of the regular expression is it's known that it has false positives, false negatives because it -- you have to make some sort of determination on the flexibility of the regular expression.

The determination I made was based on balancing the false positives and the false negatives, kind of a goldilocks-style model. But this has the advantage of being very transparent in

66

what it's capturing. So you know, the regular expression, it exists there in a few lines of code. And you can -- you know, in this case it's computer code. But I could translate this exactly into a paragraph that articulates exactly the sorts of files that it would return.

So we know precisely what the regular expression is giving. It's very transparent. It's also rep -- replicable. So in this case, regular expression -- Mr. Dennis had access to all the code. He could run the regular expression. He would get the exact same answer.

I suspect that if I were to have research associates review and taxonomize the files I see in his opening report, they would come to different conclusions. I suspect if Mr. Dennis sent those same files -- the 6,847 in the case of Exhibit 5 -- back to his own associates to categorize, that they would come up with different categorizations.

So in this case I favored a method that was transparent and replicable.

Q Okay. I just want to break that down. I appreciate the response. I want to make sure I have what you were saying.

You didn't consider the Claude refinement

67

to be providing a source of ground truth?

A That's correct. I used Claude to propose files that might be characterize -- might be examples of false negatives. And kind of as suggestions for me, as there -- there might be another common way that this sort of intention of seeking any amount or form of lyric regurgitation, there might be a common way that that is phrased that wasn't captured by my original regular expression, and so those suggestions could be used -- were -- were used to adapt the regular expression to capture that sort of expression, but in a way that didn't induce too many false positives. But it didn't serve as the ground truth.

Q And I think you said "to determine the false negative rate, you would need to apply that level of scrutiny," and I think you were discussing it in the context of Mr. Dennis's analysis and Mr. Lustig -- or sorry -- Dr. Lustig.

So do I have that right, that to determine the false -- the actual false negative rate, you would need to manually review all records in the -- in the valid, adjusted sample?

A Are you asking if I wanted to get the false negative rate for, like, this last query to

68

Question 1, Version 2, yes, it -- like, getting the precise false negative rate, you would need two things. You would need to run it over all of the files, and then you would need to have a source of the ground truth. And to the extent that you don't have either of those, it's an estimated error rate.

And you know, probably the strongest form of -- if you want to find a classifier that has the lowest error rate, I would say it's an expert in music lyrics. And -- and any -- so you're -- you know, you need that.

And for the first part, you know, running it over -- if you have to estimate this on -- if you don't have Mr. Lustig or another expert look over the 2.5 million files, you need to subsample those 2.5 million files. And the fact of the matter is, this is a very rare type of phenomenon, and that has its own challenges with estimations. So there are actually just -- just in trying to -- so let me back up.

The precise provision of the false negative rate here is extremely challenging because you need to run it on all 2.5 million, roughly, in the adjusted sample, and you need really the ground truth. And it -- but in reality, you have estimates

69

on both sides. Both are subject to error.
That sampling error, if you were to have Mr. Lustig say "a low-error classifier run on a small set of data," now you're inducing error of the first form, which now you had to estimate this error rate using a small sample.
And error rates that are that low are ex -- more difficult to estimate. And this is -- this is kind of illustrated, actually, in Mr. Dennis's -- do you happen to have Mr. Dennis's reply report?
Q Reply or his -- his rebuttal to you or his --
A His reply report.
Q I don't believe I have that.
A Okay. But I'll try to capture it to the best of my ability.
This was in his assessment of the sampling variability of his own error calculation. And this was, you know, in reply to his own revision to critique. And he comes up with an error rate of -- sorry.
A correct rate of 98 percent, which is the same error rate of 2 percent. And he provides a confidence interval on that error rate. The

70

confidence interval that he used appears to be a Wald confidence interval, which is used if you have -- trying to estimate an error rate that's closer to 50 percent.
You know, error rates, they can go anywhere between zero and 100 percent. And type of confidence interval behaves really well in proportions, percents that are anywhere in kind of the middle of the range. As you start getting towards very rare or very common events, to estimate the relative error is actually quite difficult. You need to use different methods, and essentially the uncertainty becomes greater.
So the -- he used the incorrect form of confidence -- it appears that he used the incorrect form of a confidence interval. And how you use the correct one, his error bars would have been larger.
So this is just an illustration of the fact that, you know, getting a precise false negative rate is exceedingly challenging, estimating it. In this particular circumstance it is exceedingly difficult because of two components. One is that getting the ground truth -- the ground truth doesn't exist. You're estimating it.
And the second is you're trying to

71

estimate an exceedingly rare event, which has a lot of sampling variability associated with it. So you're replacing -- you know, you're trying to quantify an error rate, but it's an error rate of an error rate of an error rate.
THE COURT STENOGRAPHER: Wait. I need you slow down, please. I didn't want to interrupt. Go ahead.
A So if you're trying to --
THE COURT STENOGRAPHER: I got that. It's okay.
A -- then it would be an error rate of an error rate of an error rate.
Q (BY ATTORNEY HOWELL) For Question 1, what would you consider to be an acceptable false positive rate?
ATTORNEY PATEL: Objection, vague.
A Question 1, what would -- so in here, acceptable -- could you define what you mean by "acceptable"?
Q (BY ATTORNEY HOWELL) Sure. You're offering an opinion that -- regarding the percentage of records that reflect a request for lyrics, correct?
A Um-hum.

72

Q And what false positive rate within those records that you identify would you be comfortable with and still be able to express your opinion that that number reflects the number of requests for lyrics in that -- in that population?
A Right. So the -- the -- for -- the strength of this approach -- when you're building a classifier and you -- like most classifiers, it has errors in both directions.
To get a good estimate of what that actual proportion is, the prevalence in this case, the criteria is not that the -- any one of the error rates is some particular magnitude, but rather, that the number of false positives is equivalent to the false negatives.
And that was the determination that was -- that was the -- what I was seeking.
Q Okay. So you were looking for -- was it a one-to-one proportion of false negatives to false positives?
A In this setting with a classifier like this, if you have an equal number of false negatives and false positives, then the proportion that you get from this will be a good estimate of the -- the true prevalence.

201

A I recall that if -- well, in a few different areas of both his opening report and his reply report, he provides extrapolations. I think extrapolations -- there are several forms of extrapolation.

One form of extrapolation that he does repeatedly is in his description of the goals of his analysis, he persistently says he is seeking -- I don't have his report, so I can't know the exact phrasing -- phrasing that he used, but something like "the proportion of Claude users that," and then he cites one of the taxonomies -- one of the categories, when, in fact, he -- he is seeking that, and yet he's not looking at records of Claude users.

He's looking at "of Claude users of files that have been filtered through counsel and then provided to him."

That's the -- kind of the reference class for his whole analysis is this subset of the files. And then he's repeatedly extrapolating to this more general sense of Claude users. And that's one form of extrapolation that's unwarranted.

The second form of extrapolation that's unwarranted is outside of the time period of the record universe, this is in particular thinking about his sample song records analysis. And the opinions that he gives on how those can be extrapolated forward in time is -- is unfounded.

And finally, within the record period within the five million -- from the five million, I should say, to the record universe within the record period, he makes extrapolations, and my critiques of those are as what's stated here.

Q Okay. What's your understanding of the LogID field?

A My understanding of the LogID field is that it is -- is intended to be a unique identifier, and specifically a unique identifier of a record that's created on Anthropic's servers -- servers after the filing of a user prompt.

So in theory there should be a one-to-one matching between somebody querying Claude; Claude replying. There should be one record. That should have a unique LogID field.

Q But you recognize that there were certain records in the sample that had the same LogID value?

A That's correct. There were a good number of records in the five million sample that both shared a LogID value and also had identical text in the extracted text field.

203

Q Okay. And I believe your initial understanding was that this LogID duplication issue was specific to Anthropic employee records?

A That's correct. And it was later -- I can't remember at what phase of the analysis, but while determining how many of these files there were that shared, the text is the exact same; the LogID field is the same.

At some point I looked at the -- the email field -- the "Email Address CF" -- and found that there are Anthropic emails, but there are also -- I can't remember if the field was registered as "missing" or "redacted," but ones that were not clearly evidently employee records --

Q And you --

A -- records that had an email address that corresponds to an Anthropic employee.

Q Okay. And you discovered that after submitting your opening report, correct?

A I believe that's true.

Q Okay. Are you offering any opinion in this case on fair use?

A Fair use is a legal notion. I have no particular legal expertise here, but I understand that the opinions that I offer here around, you know, the characteristics of these text prompts, that they have bearing on questions of fair use.

Q Are you offering any opinion on damages?

A It's a similar answer in that I have no particular expertise on damages. I know the numbers that I have used here have informed damages reports, so they have -- they certainly have -- have bearings on -- on those questions.

Q And are you offering any opinion here on whether Anthropic committed copyright infringement?

A Copyright infringement, too, is a legal term, and I have no particular legal expertise, so my opinion is confined to the reports here and the characteristics of the datasets that were provided to me and the answers to the research questions.

And I understand they might -- they might have some bearing on questions of copyright infringement.

Q Okay. Your reports include a List of Materials Considered.

A Um-hum.

Q Do you recall that?

Are there any materials you relied on in forming your opinions that are not reflected in those Lists of Materials Considered?

221

identifier of a file, and yet we see examples where we have two files.

The LogID is the exact same. The text content is the exact same. So there was some other process that happened at some point in the kind of data storage data collection process that had another source of data duplication.

Q (BY ATTORNEY PATEL) What does the appearance of multiple copies of prompt output pairs in a sample tell you about whether the record universe includes at least one separate record for every prompt and output submitted to Claude?

ATTORNEY HOWELL: Objection, form.

A So based on my understanding of the court order, the five million sample that I looked at was a simple random sample of 2.5 million from the presuit -- from the record universe in the presuit period and 2.5 million from the record universe in the post-suit time period.

And if it's a simple random sample, then you would expect that it should be representative of the population, the superset from which it was drawn.

So in my sample I observed that records could be -- prompts could be duplicated by any of

222

those three mechanisms -- the LogID duplication, the title record duplication, and the nesting process. So it's -- so it's -- based on that understanding, those all -- each one of those three types of files should exist in the record universe as well.

And kind of my understanding of the record universe also is built by just my understanding of the way that LLMs work, that they, at least in this time period, had this stateless characteristic where every prompt that was sent included all of the previous prompts.

So I would just expect that any database that they have that's storing all these interactions should be storing them in this nested structure.

ATTORNEY PATEL: Okay. No further questions from me.

ATTORNEY HOWELL: No questions from me. Thank you, Dr. Bray.

THE VIDEOGRAPHER: This -- stand by. This ends the videotaped deposition of Dr. Andrew Bray. The time is 4:10 p.m. We are going off the record.

(TIME NOTED: 4:10 p.m.)

223

I, ANDREW P. BRAY, Ph.D., do hereby declare under penalty of perjury that I have read the foregoing transcript; that I have made any corrections as appear noted, in ink, initialed by me, or attached hereto; that my testimony as contained herein, as corrected, is true and correct.

EXECUTED this_____ day of_____, 20____, at_____, _____.
(City) (State)

_____
ANDREW P. BRAY, Ph.D.

224

I, MARY J. GOFF, CSR No. 13427, Certified Shorthand Reporter of the State of California, certify;

That the foregoing proceedings were taken before me at the time and place herein set forth, at which time the witness declared under penalty of perjury; that the testimony of the witness and all objections made at the time of the examination were recorded stenographically by me and were thereafter transcribed under my direction and supervision; that the foregoing is a full, true, and correct transcript of my shorthand notes so taken and of the testimony so given;

That before completion of the deposition, review of the transcript (XX) was ( ) was not requested: ( ) that the witness has failed or refused to approve the transcript.

I further certify that I am not financially interested in the action, and I am not a relative or employee of any attorney of the parties, nor of any of the parties.

I declare under penalty of perjury under the laws of California that the foregoing is true and correct, dated this 26th day of February, 2026.

Mary J. Goff
_____
MARY J. GOFF

Dr. Andrew Bray Errata Sheet

Concord v. Anthropic

Conducted on February 24, 2026

Page: Line	From	To
9:3-5	but there – the contributions of WilmerHale in terms of drafting the reports were at my direction	but the contributions WilmerHale made to drafting the reports were at my direction
9:15-19	I understand that the Plaintiffs are – I believe that there's been copyright infringement on lyrics at a few different phases of Anthropic's product Claude and that Anthropic disputes that characterization	I understand that the Plaintiffs allege copyright infringement of lyrics at a few different phases of Anthropic's product Claude and that Anthropic disputes that characterization
12:1-2	No other areas beyond the interpretation analysis of data	No other areas beyond the interpretation and analysis of data
12:19	linguists	linguistics
12:21	linguists	linguistics
13:24-14:1	There on other areas of statistics…and classification models, models for data	There are on other areas of statistics…and classification models and other models for data
22:20-21	annual review	manual review
47:19-20	I read through a reduced spot-check	I read through a reduced sample as a spot-check
61:15	wanting to just the regular expression	wanting to adjust the regular expression
76:2	common false positive rate	common source of false positives
156:22	We already the	We already know the first
214:16	What I do access to is the -- the times at which	What I do have access to are the times at which

I, **Dr. Andrew Bray,** read the foregoing deposition and hereby state that the foregoing is true and correct, except as noted herein.



Dr. Andrew Bray

3/30/2026

Date