

BEFORE THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA, et al., .
Plaintiffs, . Case Number 20-cv-3010
vs. .
GOOGLE LLC, . Washington, D.C.
Defendant. . September 20, 2023
1:32 p.m.
- - - - -

TRANSCRIPT OF BENCH TRIAL, DAY 7
(AFTERNOON SESSION)
BEFORE THE HONORABLE AMIT P. MEHTA
UNITED STATES DISTRICT JUDGE

APPEARANCES:

For DOJ Plaintiffs: KENNETH DINTZER, ESQ.
United States Department of Justice
1100 L Street Northwest
Washington, D.C. 20005

JOSHUA HAFENBRACK, ESQ.
ERIN MURDOCK-PACK, ESQ.
United States Department of Justice
450 Fifth Street Northwest
Washington, D.C. 20001

DAVID DAHLQUIST, ESQ.
United States Department of Justice
209 South LaSalle Street
Suite 600
Chicago, Illinois 60604

-- continued --

1 APPEARANCES (CONTINUED):

2 For Plaintiffs State of
3 Colorado and State of
4 Nebraska:

WILLIAM F. CAVANAUGH, JR., ESQ.
Patterson Belknap Webb & Tyler LLP
1133 Avenue of the Americas
Suite 2200
New York, New York 10036

6 For the Defendant:

JOHN SCHMIDTLEIN, ESQ.
EDWARD BENNETT, ESQ.
KENNETH SMURZYNSKI, ESQ.
Williams & Connolly LLP
680 Maine Avenue Southwest
Washington, D.C. 20024

MICHAEL SOMMER, ESQ.
Wilson Sonsini Goodrich & Rosati
1301 Avenue of the Americas
40th Floor
New York, New York 10019

13 Official Court Reporter:

SARA A. WICK, RPR, CRR
333 Constitution Avenue Northwest
Room 4704-B
Washington, D.C. 20001
202-354-3284

16 Proceedings recorded by stenotype shorthand.
17 Transcript produced by computer-aided transcription.

C O N T E N T S

TESTIMONY

MICHAEL ROSZAK	Cross-Examination.....	1735
	Redirect Examination.....	1741
	Recross-Examination.....	1744
ERIC LEHMAN	Direct Examination.....	1748
	Cross-Examination.....	1828

EXHIBITS RECEIVED

UPX38.....	1857
------------	------

P R O C E E D I N G S

(Call to order of the court.)

(The following occurred under [REDACTED])

THE COURT: All right, Mr. Bennett.

MR. BENNETT: Thank you, Your Honor.

MICHAEL ROSZAK, WITNESS FOR THE PLAINTIFFS, RESUMED STAND

CROSS-EXAMINATION (Continued)

BY MR. BENNETT:

Q. Thank you for coming back from lunch. I have a few more minutes of questions for you.

If we could pull up UPX Exhibit 148 and go to Bates Number 5826.

I want to return to this exhibit you were shown this morning, Mr. Roszak. I think when you were asked questions about the Mozilla and the Google Maps for iOS scenarios, you were asked a question about whether the data was observed, and I think you said it wasn't observed.

What did you mean by that?

A. I'm trying to remember the context.

Q. I think the context was about whether this was data that Google observed during those instances or measured during those -- is this data that is the -- that precisely captures what happened with usage with the Firefox change or with the Apple iOS change?

A. I mean, I wouldn't say it's precise. A lot of it is trying

1 to look at trend lines and infer what might be happening based
2 on the trends that we're seeing. So I'm not -- I'm not trying
3 to remember the specifics, but a lot of it is inferred based on
4 what we are seeing in trend lines like this.

5 Q. Thank you. That's helpful.

6 So when you were using -- and let's focus now on the iOS
7 maps information. When you were using the information relating
8 to the change of iOS maps from Google being the embedded mapping
9 system on iOS to Google -- to Apple launching its own map app,
10 it used that data to make estimates regarding search recovery
11 and other later scenarios.

12 Did you adjust your assumptions to account for the fact
13 that when Apple launched its map, its mapping application and
14 turned away from Google, that Google did not have an app. for
15 iOS, a mapping app. for iOS?

16 A. No.

17 Q. And I think it was established that there was a few-month
18 period when Google did not have an iOS mapping app.

19 Is that your recollection?

20 A. That's my understanding.

21 Q. And the effect of that period when users were unable to, on
22 a new iPhone or on an iPhone that took the update, unable to
23 access Google Maps, did the effect on users factor into your
24 later calculations or assumptions, rather, regarding estimated
25 recovery?

1 A. No.

2 Q. Now, do you know whether or not it's the case on Safari
3 that Apple Maps is hard coded as the default mapping app.?

4 A. I am not certain.

5 Q. Do you know whether or not a user of an iOS phone can
6 switch the default mapping app. in iOS -- I'm sorry, in Safari?

7 A. I am not certain.

8 Q. Okay. And I take it, then, in doing your later estimates,
9 you didn't adjust your assumptions for the fact that there's no
10 way to change the default mapping program for Safari?

11 A. No, we did not adjust for that.

12 Q. And when you used the Apple Maps scenario in making later
13 estimates, did you adjust your assumptions for different ways
14 that users might use a mapping program as opposed to a search
15 engine?

16 A. No.

17 Q. Let's go to Exhibit 462, and if we could pull up the bottom
18 of this e-mail, the e-mail from Evan Sidarto.

19 Mr. Roszak, this is an exhibit you were shown earlier about
20 RPM in Japan. Do you remember seeing this exhibit earlier?

21 A. Yes.

22 Q. And you were focused on the top e-mail. I want to look at
23 the bottom e-mail. First of all, let me just establish some
24 basics here.

25 Who is Mr. Sidarto?

1 A. He is the finance lead for the Asia-Pacific region.

2 Q. And you had asked him a question that we heard earlier, and
3 he came back with this response about what could account for the
4 different RPMs. And he starts in his e-mail in the second -- or
5 third sentence, "I don't think any of them are fair comparisons
6 for the U.S., UK, or Australia given the language differences."

7 Is that one of the factors that the Asia-Pacific finance
8 lead told you was affecting RPMs in Japan?

9 A. Yes. He says he doesn't think it's a fair comparison,
10 given that.

11 Q. And he goes on to list several other bullet points about
12 mSearch in Japan is the most mature in the world.

13 What does that mean?

14 A. It looks like that is saying mobile search, and based on
15 the parenthetical after that, I think he's saying in Japan the
16 percent of mobile search out of all searches is the highest in
17 that market.

18 Q. And then the next bullet --

19 MR. HAFENBRACK: Your Honor, I would just note that we
20 did this document in open session. So I'm not sure it's
21 appropriate to do the cross in close.

22 THE COURT: Okay.

23 MR. BENNETT: Two more questions on it. I can do
24 those in open, or I can do them now.

25 THE COURT: Why don't we just wait until the session

1 is reopened. I know it's more efficient to do it, but let's
2 just stay consistent.

3 Thank you for the reminder.

4 BY MR. BENNETT:

5 Q. Let's go to Exhibit 474 -- I'm sorry, 674. I can't read my
6 writing. And I want to go here specifically to Bates ending in
7 914.

8 914 is the spice view that you saw earlier. I'm not going
9 to go all the way back through it, but I do want to focus you on
10 the table in the lower left where there's some math being done.

11 First, in doing the estimates about how Microsoft might
12 perform or Bing might perform as the default search engine on
13 Apple products, did you consider any uplift, like reputational
14 uplift that Microsoft might enjoy by displacing Google as the
15 default on Apple devices?

16 A. No.

17 Q. And did you consider any -- conversely, any reputational
18 harm that Google might suffer by losing that position to Bing?

19 A. No.

20 Q. And when you -- there's reference to clawback in here in a
21 couple of places, once with Bookmark and once with additional
22 clawback below that.

23 In estimating those clawbacks, did you take into account
24 any changes that Apple might make to its system to favor a Bing
25 Apple experience over a Google Apple experience for users who

1 might want -- might be interested in switching back to Google?

2 A. No.

3 Q. Now, let's finally go to Bates -- same exhibit, Bates 921.

4 First of all, you were shown a couple of decks for NYC
5 scenarios. Were there several drafts of that sent around?

6 A. This would have evolved across time. There were several
7 drafts of the materials.

8 Q. This is one, I think the government pointed out, is later
9 in time than one of the other ones they showed you. And I want
10 to focus on the mobile defaults where there is a statement the
11 government asked you about.

12 "Mobile defaults: Defaults have more prominence in mobile
13 due to screen size and UI." The government asked you several
14 questions about that.

15 And my question is whether you're aware of any analysis or
16 surveys or anything that was conducted to make that hypothesis
17 that shows up here.

18 A. I am not aware of any surveys or data or anything that
19 underpin that statement.

20 Q. Do you know where it came from?

21 A. I would think it probably came from like a hypothesis from
22 conversations, I guess, of, you know, one assumption to make
23 here. I don't -- as far as I know, I don't know if there was
24 anything that it was based on other than just a hypothesis.

25 MR. BENNETT: Thank you. That's all I have for the

1 closed session, Your Honor.

2 THE COURT: What was the last exhibit again, please.

3 MR. BENNETT: We were in 674. I said it incorrectly
4 the first time. I apologize for that.

5 THE COURT: All right. Okay. Before we do that, any
6 redirect in the closed session?

7 MR. HAFENBRACK: Yes, Your Honor.

8 REDIRECT EXAMINATION

9 BY MR. HAFENBRACK:

10 Q. If we could pull UPX38 back up.

11 Mr. Roszak, did you give the speech that these notes are
12 notes for?

13 A. I expect --

14 THE COURT: This a little bit out of the scope.

15 He did ask about this. Pardon me. Go ahead and ask your
16 question. Sorry.

17 BY MR. HAFENBRACK:

18 Q. My question is whether -- UPX38 reflects your notes for a
19 speech you were going to give at a Google training; correct?

20 A. For a presentation, yes.

21 Q. And my question is whether you actually gave the
22 presentation that's listed here.

23 A. I expect I would have done a presentation. I don't know if
24 I would have said this exactly. These were notes. But I expect
25 I would have given some presentation.

1 Q. Do you recall whether you used these notes as sort of your
2 outline for the presentation?

3 A. I think I probably would have used them as the outline. I
4 don't know if I would have followed them. I actually try in
5 presentations to not have notes.

6 Q. And when you were giving the presentation, were the other
7 Google employees that were a part of the training, were they the
8 audience for the presentation?

9 A. I don't know that there really was an audience. It was
10 much more around like coaching on how to give presentations. I
11 didn't really view it as audience or anything else, given it was
12 just like -- I mean, it really wasn't audience so much as, you
13 know, how to use some of the tips that they had on
14 presentations.

15 Q. Okay. If I could direct your attention to the fifth bullet
16 point that said "we had a few jolts across time," and then you
17 provide two examples of jolts. One is iOS 7.

18 And that's when Apple changed the Safari search bar
19 configuration; right?

20 MR. BENNETT: Objection. Now we are beyond the scope,
21 Your Honor. I asked about the first full bullets, and that was
22 it.

23 THE COURT: Let me hear the question first, and then I
24 can decide. Go ahead.

25 BY MR. HAFENBRACK:

1 Q. My question is whether the iOS 7 example that you provide
2 was an example where Apple changed the configuration of the
3 Safari browser.

4 A. Can I answer? Sorry.

5 Yes, I believe so.

6 Q. And the second example you provide here is Mozilla
7 switching the default to Yahoo!; right?

8 A. Yes.

9 Q. And those are just facts that you included in your
10 presentation; right?

11 A. Yes.

12 Q. And you're aware of those facts based on your work at
13 Google; right?

14 A. Yes.

15 Q. Okay. Your counsel --

16 THE COURT: I will overrule the objection on scope.
17 Go ahead.

18 MR. BENNETT: Thank you, Your Honor.

19 BY MR. HAFENBRACK:

20 Q. If you could pull up UPX85.

21 And you can turn to that in your binder if you want, or you
22 can look at it on the screen, whatever you prefer, Mr. Roszak.

23 And you see the date here of your e-mail to Ms. Porat,
24 October 2015; right?

25 A. Yes.

1 Q. And then if you go to the third page, back to that same
2 bullet we were talking about earlier, maintaining the Safari
3 default distribution is paramount, it describes the 30 percent
4 revenue recovery as the dooms-day scenario.

5 Do you see that?

6 A. Yes.

7 Q. And subsequent to this e-mail in October 2015, Google
8 started anchoring on the 30 percent revenue estimate for Apple
9 iPhones; correct?

10 A. We started using that in our scenarios, yes.

11 MR. HAFENBRACK: Nothing further, Your Honor.

12 THE COURT: Okay. Let's go ahead and open the
13 courtroom up, and we can finish up with Mr. Roszak. We're just
14 going to need a couple of minutes to make sure we get the media
15 room line reconnected.

16 MR. BENNETT: I'm embarrassed to go through this.
17 I've got about 30 seconds.

18 (End of sealed proceedings.)

19 THE COURT: Mr. Douyon went down to let everybody know
20 the courtroom was reopen. It's going to take about five minutes
21 to reconnect the line. Instead of waiting, why don't we plow
22 forward.

23 MR. BENNETT: Thank you, Your Honor.

24 RECROSS-EXAMINATION

25 BY MR. BENNETT:

1 Q. Exhibit 462, it's the Japan RPM e-mail you were asked some
2 questions about. And there were questions posed to two of your
3 colleagues who, I understand from prior questioning, are the
4 head of finance for Google in Japan and the head of finance for
5 Google in Asia; is that correct?

6 A. Asia-Pacific, yes.

7 Q. Asia-Pacific, I'm sorry.

8 And so with those duties, are those people you would turn
9 to if you wanted to understand the dynamics of competition for
10 search advertising in the United States?

11 A. No, those are not individuals I would reach out to on that.

12 MR. BENNETT: I have nothing further, Your Honor.

13 THE COURT: Just for the interest of completeness on
14 the public record, will you re-ask the questions you asked him
15 in closed session and elicit that information?

16 MR. BENNETT: Oh, sure, Your Honor. I apologize for
17 that.

18 BY MR. BENNETT:

19 Q. So I had previously asked you a question about who
20 Mr. Sidarto is.

21 Can you tell us who Mr. Sidarto is?

22 A. He is the Google finance lead for Asia-Pacific.

23 Q. And at the bottom of Exhibit -- I guess we need to call up
24 the exhibit as well, 462, and we're looking at the bottom
25 e-mail.

1 At the bottom of 462, are there several bullet points that
2 Mr. Sidarto provided you in response to your question about RPMs
3 in Japan?

4 A. Yes.

5 MR. BENNETT: That's all I have, Your Honor. Thank
6 you.

7 THE COURT: Okay. Any redirect?

8 MR. HAFENBRACK: No, Your Honor.

9 THE COURT: Mr. Roszak, thank you very much for your
10 time and your testimony, and safe travels home.

11 THE WITNESS: Thank you.

12 THE COURT: Okay. Are plaintiffs ready with their
13 next witness?

14 MR. HAFENBRACK: Your Honor, if I may, at this time
15 the United States moves into evidence UPX38 under Rule
16 801(d) (2) (D).

17 MR. BENNETT: Objection, Your Honor.

18 THE COURT: Let's hold off. I actually didn't bring
19 my materials with me. So why don't we take that up at the start
20 of the next break.

21 MR. HAFENBRACK: Sure, Your Honor.

22 THE COURT: Or right after the next break, I should
23 say.

24 MR. SOMMER: Judge, there was one exhibit yesterday
25 that Your Honor received provisionally but DOJ wanted to

1 consider it overnight. It's DXD03.005.

2 MR. DAHLQUIST: Yes, Your Honor, and to the extent
3 that's offered as a demonstrative, I have no objection. But to
4 the extent it's being offered substantively, we would.

5 MR. SOMMER: We are offering it pursuant -- do you
6 want a copy?

7 THE COURT: Yes, please.

8 MR. SOMMER: We are offering it pursuant to 1006, Your
9 Honor. All of the underlying exhibits that form the data are
10 unobjected to. It's simply a compilation.

11 THE COURT: And who prepared it?

12 MR. SOMMER: Dr. Israel, one of our excerpts.

13 THE COURT: Any objection to it being admitted as a
14 1006?

15 MR. DAHLQUIST: I think they have to ask Dr. Israel
16 that in order to get his view on it.

17 THE COURT: I'm asking you whether you -- well, if
18 you're not prepared to agree to its admissibility, are you
19 reserving the opportunity to question Dr. Israel about his
20 preparation of this?

21 MR. DAHLQUIST: Yes, Your Honor.

22 THE COURT: Okay. So let's just wait, and we can ask
23 Dr. Israel how he created this chart and go from there.

24 MR. SOMMER: Fine.

25 MR. DAHLQUIST: Thank you, Your Honor.

1 MS. MURDOCK-PARK: Good afternoon, Your Honor. Erin
2 Murdock-Park on behalf of the United States. We call Eric
3 Lehman as our next witness.

4 ERIC LEHMAN, WITNESS FOR THE GOVERNMENT, SWORN

5 MS. MURDOCK-PARK: Because Google has asserted
6 confidentiality over some of the subjects of Dr. Lehman's
7 testimony, we expect to conduct both open and closed sessions
8 today.

9 THE COURT: Okay.

10 MS. MURDOCK-PARK: I hope to end the public session
11 around our next break time.

12 DIRECT EXAMINATION

13 BY MS. MURDOCK-PARK:

14 Q. Good morning, Dr. Lehman. My name is Erin Murdock-Park. I
15 represent the United States.

16 You were a distinguished software engineer at Google;
17 correct?

18 THE COURT: Could we just ask the witness to state his
19 name and spell it.

20 MS. MURDOCK-PARK: My apologies.

21 BY MS. MURDOCK-PARK:

22 Q. Mr. Lehman, could you please state and spell your name for
23 the record.

24 A. Lehman, L-e-h-m-a-n.

25 THE COURT: Thank you, sir.

1 BY MS. MURDOCK-PARK:

2 Q. And Dr. Lehman, you were a distinguished software engineer
3 at Google?

4 A. At the end of my time at Google, yes.

5 Q. And you left Google in November 2022?

6 A. Yes.

7 Q. Why did you leave Google?

8 A. I found that the work was becoming very demanding, and I
9 couldn't balance work with my family responsibilities.

10 Q. What is your current job?

11 A. I am a stay-at-home father.

12 Q. Are you represented by counsel today?

13 A. Yes.

14 Q. Whom?

15 A. Ken, I guess.

16 MR. SMURZYNSKI: Smurzynski. It's not easy.

17 MS. MURDOCK-PARK: Thank you.

18 BY MS. MURDOCK-PARK:

19 Q. Do you have a pension from Google, Dr. Lehman?

20 A. A pension? No.

21 Q. Do you own Google stock?

22 A. I don't myself own Google stock. However, my wife is an
23 employee of YouTube, which is a part of Google, and she does
24 periodically get stock grants.

25 Q. Do you have any financial interest in the outcome of this

1 case?

2 A. Only to the extent -- only via my wife's employment, I
3 guess.

4 Q. Now, the questions I'll be asking you today will revolve
5 around the time that you spent working at Google. Okay?

6 A. Okay.

7 Q. Let's talk briefly about your background. You received a
8 bachelor's, master's, and Ph.D. from MIT; is that correct?

9 A. That's right.

10 Q. And you also did a post-doctoral at MIT?

11 A. That's right.

12 Q. Then you worked at Google for about 18 years?

13 A. I think a little over 17.

14 Q. You always worked in search quality and ranking?

15 A. Yes, that's right. I always worked in search quality on
16 ranking with some sort of peripheral things.

17 Q. Thank you. At one time, you managed a team of about 50
18 people who worked in search quality; correct?

19 A. That's right. Maybe a year or so before I left, I stepped
20 down from management, but the largest group I managed was just
21 before that and was about 50, yes.

22 THE COURT: What was your Ph.D. in?

23 THE WITNESS: Theoretical computer science, so
24 combinatorial and probabilistic algorithms.

25 THE COURT: Thank you.

1 BY MS. MURDOCK-PARK:

2 Q. Pandu Nayak was your direct supervisor for a time when you
3 were at supervisor?

4 A. Yeah, for all of my later years at Google, he was my
5 supervisor.

6 Q. You were familiar with Google's processes for ranking web
7 results; correct?

8 A. I'm generally familiar, yes.

9 Q. And you're familiar with Google's efforts to refine and
10 improve its search algorithms?

11 A. Yes, I'm generally quite familiar.

12 Q. You're also familiar with Google's use of user interaction
13 data; right?

14 A. Fairly familiar. I didn't sit on the team that is most
15 heavily engaged with user interaction data, but I had many
16 connections to it during my time.

17 Q. Can you briefly explain for the Court what user interaction
18 data is with respect to search quality?

19 A. Yeah, so here's the idea. The biggest problem in search
20 is, given a query and a web page, is that web page relevant to
21 the query. A person can figure that out pretty easy. They just
22 read the web page and think about it. But a search engine runs
23 on computers, and they couldn't read.

24 So there's sort of a trick to get around that, which is you
25 show web pages to people and then see their reaction, and that

1 can give you some sort of hint about whether or not the web page
2 is relevant to the query. So you do a search. You see five
3 search results. The person can read the query, can reach the
4 search results, and then they might click on one of those
5 results or skip over one of them, user interaction. And that
6 gives you some hint about what it is the person read.

7 So I think that's user interaction data you're talking
8 about.

9 Q. Yes. So briefly, so user interaction data provides users'
10 feedback about search results? Is that a fair summarization?

11 A. Did you say user interaction data provides users?

12 Q. Provides feedback from the users about their search
13 results.

14 A. I think that's fair, yeah.

15 Q. And user interaction data is Google's observation of the
16 user's engagement with the search results?

17 A. Yes. So if someone does a search, then they might look at
18 the results for a little bit. They might click on something,
19 back up, click on something else, maybe swipe through a
20 carousel. Usually, those events is logged and time-stamped and
21 recorded, and that becomes a body of user interaction data.

22 Q. So Google sometimes uses the term "clicks" as a shorthand
23 for user interaction data; right?

24 A. Sometimes, yeah, yeah. I mean, people who work on this
25 stuff really probably make the fine distinctions, but we can use

1 clicks as sort of a shorthand for user interaction data.

2 Q. And you mentioned that another team was more heavily used
3 or -- strike that.

4 Which team at Google is most heavily engaged with user
5 interaction data?

6 A. It's a team that has sort of gone by various names:
7 Navboost, Glue, Tetris, things like that. It's all kind of the
8 same group of people with their role sort of evolving a little
9 bit over time.

10 Q. Thank you. If I use "user data" instead of "user
11 interaction data," will you understand what I mean, or would you
12 prefer I use the whole term?

13 A. I think we can agree what user data is.

14 Q. Okay. So let's talk a little bit more about how Google
15 views the use of user data in its search systems. And I would
16 like to turn your attention to a slide deck that you authored in
17 2017 titled "Google is magical." This is UPX228. And this is
18 admitted, and there are no redactions on this document.

19 Dr. Lehman, you drafted UPX228; correct?

20 MR. SMURZYNSKI: Your Honor, could the witness be
21 presented with the document before he is questioned about it?

22 MS. MURDOCK-PARK: Oh, my apologies. If may I
23 approach.

24 THE WITNESS: Might you be able to tell me the number
25 one more time?

1 BY MS. MURDOCK-PARK:

2 Q. 228. You will notice in the binder, Dr. Lehman, there is
3 an "open" section, and that's what we will be talking about
4 publicly. And there is a "closed" section, and that's what we
5 will be discussing in the sealed session.

6 So the first half of the binder will have documents we're
7 discussing now. I think it's the sixth tab in.

8 A. I see. It says "closed"?

9 Q. Yes. We're in the "open" section right now at UPX228.

10 A. Okay.

11 Q. Do you have the document in front of you, Dr. Lehman?

12 A. Yes.

13 Q. Okay. So did you draft the presentation "Google is
14 magical"?

15 A. Yes.

16 Q. And you titled the slide deck "Google is magical" because
17 Google gets search results right even when that doesn't seem
18 possible; correct?

19 A. There's a somewhat complicated story behind this. I'm not
20 sure that that quite captures the reason for that.

21 Q. Okay. We'll get into a little bit more why you drafted the
22 deck.

23 A. Okay.

24 Q. But you presented UPX228 to other Google employees?

25 A. I just want to check, because I did probably variations on

1 this presentation several times. But yes, yes, I did.

2 Q. Let's turn to slide 7, which is Bates ending 9499, and this
3 is a graphic titled "this is not how search works."

4 Do you see that?

5 A. Yes.

6 Q. Okay. And the graphic on slide 7 has one set of arrows
7 going from left to right. The keyboard on the left that
8 has "scoring" over it, that represents Google?

9 A. That's right.

10 Q. And the "results" arrow pointing from Google to an artist
11 pallet that's labeled "UX," does "UX" stand for user experience?

12 A. That's right.

13 Q. And the second arrow labeled "SERP" going from the UX to a
14 smily person holding a phone, "SERP" stands for search engine
15 results page; right?

16 A. Search results page, yeah.

17 Q. Okay. And so a SERP includes those blue links, the web
18 results?

19 A. Yeah, blue links, image links, maybe a translation box,
20 weather maps, all kinds of things.

21 Q. Okay. And the translation box, maps, those are referred to
22 as search features; right?

23 A. Yes, those are often called search features.

24 Q. You write on slide 7 that it the graphic is not false.
25 Specifically, the last line of the user notes, "This is not

1 false, just incomplete. So incomplete that a search engine
2 built this way won't work very well. No magic."

3 Right?

4 A. Yes, that's what I wrote.

5 Q. So Google needs more than just good engineering to produce
6 this magic; right?

7 A. So at this time, for the examples that I gave in this
8 presentation, we also needed user feedback data.

9 Q. And you said you used other versions of this presentation
10 years following this presentation; right?

11 A. Sorry --

12 Q. In the years following this presentation.

13 A. I used portions of this in a subsequent presentation.

14 THE COURT: I'm sorry. I don't know if we got when he
15 first developed this and started presenting it.

16 BY MS. MURDOCK-PARK:

17 Q. You drafted this presentation in 2017; is that right,
18 Dr. Lehman?

19 A. That sounds about right to me, yes.

20 Q. If we could go to slide 10, which is Bates ending 9502.
21 It's titled "how search does work."

22 Do you see that graphic, Dr. Lehman?

23 A. Yes, I do.

24 Q. And slide 10 has the same graphic we looked at before, but
25 it includes additional information flow from the user back into

1 Google; right?

2 A. That's right, yes.

3 Q. And the user is wearing a magic hat?

4 A. Yes.

5 Q. And both arrows pointed back to Google have on magic hats
6 as well; right?

7 A. Yes.

8 Q. So the arrow going away from the user which is
9 labeled "interaction," that represents user interaction data?

10 A. Yes.

11 Q. And the "interaction" arrow points to a graphic of logs.
12 That represents Google logging user data; right?

13 A. Yes.

14 Q. And then there's a "learning" arrow going back to Google.
15 That represents Google learning from the user data that it
16 logged; right?

17 A. That's correct.

18 Q. Okay. And part of learning from user data means training
19 Google systems; correct?

20 A. So learning from user data, there are many different
21 systems that learn from user data. The word training is most
22 commonly used in connection with machine learning systems, and
23 not all the systems that Google uses to process user data
24 involve machine learning.

25 Q. Okay. But some of the systems that Google uses do train on

1 user data; right?

2 A. Some of the systems that Google uses in ranking search
3 results are machine learning systems that train on user data.

4 Q. And so in terms of training a system, would one example be
5 something like people misspelling words and Google training its
6 systems to figure out the best way to correct those
7 misspellings? That's a pretty simple example.

8 A. Yes. That's been a very fast-changing system, and I
9 don't -- I'm a little hazy on where user interaction data is --
10 or whether user interaction data is used in that system. I
11 think it is, but I'm not sure.

12 Q. The first two lines of your speaker notes read, "The key is
13 a second flow of information in the reverse direction. As
14 people interact with search, their actions teach us about the
15 world."

16 Do you see that?

17 A. Yes.

18 Q. So you were explaining the importance of user data to
19 Google; right?

20 A. Yeah, I was explaining the importance of user data to
21 Google at that time.

22 Q. And you write in your last sentence, "We log these actions,
23 and then scoring teams extract both narrow and general
24 patterns."

25 Right?

1 A. Yes.

2 Q. Okay. So user data can improve specific search results,
3 correct, with respect to a specific query?

4 A. So for each specific query, there's a scoring process for
5 all the results, and then results are ranked by decreasing
6 store. And the user data may affect the scoring of those
7 individual results and so affect how they're ranked, yes.

8 Q. And user data can also be used to improve Google's systems
9 more broadly; fair?

10 A. Which systems?

11 Q. Well, generally speaking. There's multiple systems at
12 Google; right?

13 THE COURT: I think you can be more specific. It
14 seems like "systems" is a pretty broad word.

15 MS. MURDOCK-PARK: Certainly, Your Honor.

16 BY MS. MURDOCK-PARK:

17 Q. I'm not going to get into the specifics in the public
18 session. So if this is something that we need to talk about in
19 private session or sealed session, please let me know.

20 A. Okay.

21 Q. But generally speaking, Google uses multiple systems to
22 rank search results; correct?

23 A. So the ranking of search results involves many, many
24 components. Is that answering your question?

25 Q. Would it be easier if I referred to components rather than

1 systems? Is that the term that you prefer?

2 A. Okay. So we will say the pieces of the ranking system are
3 components.

4 Q. So there's multiple ranking components; right?

5 A. Yes, there are multiple ranking components.

6 Q. And the better that Google utilizes user data, the more
7 inferences that multiple components can make; right?

8 A. Let's see. Okay. So the better -- sorry. Could you
9 repeat that one?

10 Q. Let me hopefully ask a better question.

11 Can Google make broader inferences from the more -- strike
12 that.

13 The more user queries that Google has, the more inferences
14 it can make from those queries; right?

15 A. It's changing pretty fast. So in one direction, it's
16 better to have more user data. At the same time, with
17 technology improvements, later systems in some cases require
18 much, much less user data. So it's a little bit of a
19 complicated picture.

20 Q. All right. But the later systems still do require some
21 user data; right?

22 A. Some later systems require user data, and some do not.

23 Q. I'd like to go to the next slide on UPX228, which is slide
24 11, and this is, again, the same graphic, but this time, the
25 user, Google, and all four arrows have magic hats; right?

1 A. This is 503?

2 Q. It is, yes, Bates ending 503.

3 A. Yes.

4 Q. Okay. And in your speaker notes, you write, "The source of
5 Google's magic is this two-way dialogue with users"; right?

6 A. Yes.

7 Q. And that's why you put the magic hats on every step in the
8 process?

9 A. Yeah, to show that information flows through this whole
10 loop, yes.

11 Q. And you go on to write, "With every query, we give some
12 knowledge and get a little back. Then we give some more and get
13 a little back. These bits add up. After a few hundred billion
14 rounds, we start looking pretty smart"; right?

15 A. Yes.

16 Q. Okay. So Google has dialogues with users billions of times
17 a day; correct?

18 A. I believe that's right.

19 Q. And each of these billions of times for each user, Google
20 logs user data; right?

21 A. Yes.

22 Q. The final sentence in your speaker notes reads, "This isn't
23 the only way we learn, but the most effective"; right?

24 A. Yes.

25 Q. So Google learns in other ways, like its engineers

1 developing algorithms; right?

2 A. Yes, that's one other way that Google learns, and there are
3 others.

4 Q. Okay. But when you wrote UPX228, you considered the
5 dialogue with users to be the most effective way that Google
6 learned; right?

7 A. I think when I wrote this I considered that the most
8 effective way Google learns. Obviously, the situation is
9 changing rapidly.

10 Q. So when you wrote this, you considered the most effective
11 way Google learns is the dialogue with users; right?

12 A. So at that time, yes.

13 Q. The more searches that people run on Google, the more user
14 data that Google gets; right?

15 A. So generally, that's true. As more people do searches,
16 more data Google gets. However, Google also begins dropping
17 data at some point. So there's a flow in and a flow out.

18 Q. The more user data that Google gets, the more it can use
19 that data in various -- in various ways to improve Google
20 systems -- or Google components, the ranking system; right?

21 A. Let's see. Could you ask that one --

22 Q. Sure. Let me ask it slightly clearer.

23 The more user data that Google has, the more it can use
24 that data to improve Google's ranking components?

25 A. Well, for whatever volume of data we have, of course, we

1 make copies of it, and each copy goes to a different -- goes to
2 a different ranking component. So it isn't like we need 20
3 times as much data if we have 20 components.

4 Let's see. I'm not quite sure I'm getting --

5 Q. Okay. How about I ask a different question.

6 Would you agree that the better that Google's ranking
7 components are, the more that users should run searches on
8 Google?

9 A. So the better the ranking components are, the more users
10 should run?

11 Q. Google wants more users to run searches on its website,
12 right, on google.com?

13 A. Well, I don't know about what Google wants. When I was in
14 play at Google, I wanted to help as many people as possible get
15 access to the information they wanted.

16 Q. Okay. And that would -- when you were working at Google,
17 you considered that to be people using Google services; right?

18 A. Right. So I wanted to help as many people using Google
19 Search to get access to the information that they wanted.

20 Q. And the more people that use Google Search, the more money
21 that Google makes; correct?

22 A. I'm not a finance person. I think so.

23 Q. We can go ahead and put UPX228 down, and I want to talk a
24 little bit more about the types of user data. Okay?

25 Google turns user data into signals; correct?

1 A. I guess the term "signal" is pretty ambiguous. I guess we
2 have to be a little bit more specific about what you mean
3 by "signal."

4 Q. Okay. You've heard the term "signal" when you were working
5 at Google; right?

6 A. Yes.

7 Q. Can you explain in a few sentences for the Court what a
8 signal is in the context of user data?

9 A. So I guess when I was working on search, people would
10 generally use the term "signal" to mean some sort of chunk of
11 data that's carrying some kind of meaning.

12 So I guess a signal in the context of user data -- I guess
13 that doesn't really change the meaning of signal in the context
14 of user data. It's a chunk of information that carries some
15 kind of meaning.

16 Q. Would an example of a signal be if a document contains the
17 same words as a query?

18 A. One could call that a signal, yeah.

19 Q. Okay. And then Google interprets that signal into
20 determining that the document might be relevant to the query;
21 right?

22 A. So for that particular signal, say counts of the number of
23 times each query term appear in a document, that signal or
24 collection of signals could then be used to help assess the
25 relevance of the document to the query.

1 Q. I'd like to show you UPX4, which is a Google document
2 titled "life of a click user interaction."

3 And this is admitted. There are some redactions on the
4 slide. I'm going to start at slide 2.

5 THE COURT: Sorry. Can I just back up for a moment?
6 The term "signal" I understood to be a user action that conveys
7 some type of intent to the search engine that would enable the
8 search engine to process a result.

9 And what I just heard was an example that seems to suggest
10 that the contents of a web page could signal those results.

11 Am I misunderstanding the term, the concept?

12 THE WITNESS: The word "signal" in the sense that I
13 used it while working on search at Google was fairly broad.
14 Maybe examples would help. They're not all related to user
15 interactions.

16 So for example, a signal might be how many links on the web
17 are there that point to this web page or what is our estimate of
18 the sort of authoritativeness of this page.

19 So it's a very broad, inclusive term, I would say.

20 THE COURT: Okay. I think I understand. Thank you.

21 BY MS. MURDOCK-PARK:

22 Q. Do you see UPX4 in front of you, Dr. Lehman?

23 A. Yes.

24 Q. Okay. And you're familiar with UPX4, correct, the slide
25 deck titled "life of a click (user interaction)"?

1 A. I did not produce this slide deck, and I have not read
2 through it.

3 Q. Okay. You reviewed this document at your deposition.

4 A. Okay.

5 Q. Correct? Do you recall that now?

6 A. No.

7 Q. Let me ask you some questions on slide 5, and you can tell
8 me if the graphic looks familiar to you.

9 A. Okay.

10 Q. Have you seen the graphic at slide 5? It's Bates ending
11 8265.004.

12 A. It's the blue -- oh, I see it now, yes.

13 Q. And the graphic is titled "user interaction signals"?

14 A. Yes.

15 Q. Okay. The top left of the graphic on slide 5 shows a smily
16 faced user; right?

17 A. Yes.

18 Q. And then the "queries" area going down from the user is
19 pointing to a cloud labeled "search"; right?

20 A. Yes.

21 Q. And those queries are logged by Google; right?

22 A. Maybe I could just look at this graphic for a minute.

23 Q. Of course.

24 A. I'm not sure I understand it yet.

25 Okay. I guess if the little blue guy is a person using

1 Google Search and the arrow is queries done by that person and
2 the cloud represents, I guess, like Google, then the queries
3 would be flowing to Google's systems, I think.

4 Q. Okay. And the queries flowing into Google's system are
5 logged by Google; correct?

6 A. That's right. When someone enters a query in the search
7 box, Google receives that and logs it.

8 Q. And then there's a "results" arrow towards the bottom of
9 the page that's going towards the text box on the right.

10 Do you see that?

11 A. Yes.

12 Q. And that represents Google returning search results in
13 response to the user's query?

14 A. I think so.

15 Q. Then there's an "interaction" arrow, and I'm going to point
16 to the first one at the top from the smily face to the right
17 side, and we've highlighted it on the screen.

18 It's "interactions" and has a box with some other words in it.

19 Do you see that?

20 A. Yes.

21 Q. And those are four categories of user interactions; right?

22 A. I think so, yeah.

23 Q. Okay. Well, let's talk very briefly about those four
24 categories. The first category listed is "read." A read
25 interaction measures whether the user gave the result some

1 attention; correct?

2 A. I'm not sure whether that's read or read. Is it the user
3 interaction of reading it or -- because we don't necessarily
4 know whether somebody has read something.

5 Q. Okay. Google instead refers to things as "attention";
6 correct? Like the attention that a user may have spent on a
7 website or how long?

8 A. So I believe that attention usually refers to a person
9 looking at a portion of the search results page. Because as
10 long as they're on Google, I mean, they're looking at a Google
11 web page, then we have a chance to sort of see what are they
12 doing, we can see are they looking at the top of the page or the
13 bottom or something like that.

14 So I think it's attention to the parts of the search
15 results page as opposed to some page off of Google.

16 Q. If a user clicks on a document, then there are different
17 lengths of clicks; right?

18 A. Yeah. So when someone clicks on a document, at that point
19 they've left Google, and various things could happen. They
20 could come back to Google later, after some period of time.
21 They could just go away forever and we never see them again.

22 There are also more complicated things like if people do
23 tabbed browsing. So they will do a search, get a bunch of
24 search results, decide three of them are promising, and they
25 open three tabs, and then things start to get complicated.

1 Q. You've heard the terms "short click" and "long click";
2 right?

3 A. Yes.

4 Q. Can you please briefly explain what a short click and a
5 long click are?

6 A. Yeah, so the precise definitions have varied a lot over the
7 years, but a short click would typically be one where somebody
8 clicks on a search result and then comes back to Google fairly
9 soon after.

10 Whereas, a long click would be they click on something and
11 maybe they come back in two minutes or something. I don't know
12 how we classify like if somebody clicks on a search result and
13 we just don't see them again for that day.

14 Q. Okay. Generally speaking, a short click is associated with
15 a less relevant result; right?

16 A. It's -- my understanding is that there is a correlation
17 between click duration, but it's fairly weak. There are just a
18 lot of complications there. For example, if you're just trying
19 to get the telephone number of something, you search for it.
20 You click on a search result. You get the phone number, come
21 back. You're completely happy.

22 Also, media type can have a big effect. So you do a
23 search. You click on a video. You go away. You watch it for
24 five minutes. They're saying hey, everyone subscribe to my
25 channel and all that. So that can sort of protract a click.

1 So the click duration, it's a pretty noisy signal, but I
2 think in general the thinking is long clicks are typically
3 better than short clicks.

4 Q. Let's go back to slide 5 in UPX4 and look at the second
5 category of user data, which is clicks.

6 We've talked about clicks. That's when somebody just
7 clicks on a web result; right?

8 A. Sorry. Slide 5?

9 Q. It's the slide that's up on the screen right now, slide 5
10 of UPX4.

11 A. Okay.

12 Q. And at the top where it has the "interactions" box and then
13 there's the "clicks."

14 A. You're calling it slide 5. Is it 004?

15 Q. It's UPX4, and the Bates number at the bottom is 265.004.

16 A. Got it. Okay.

17 Q. All right. The second "clicks" under "interactions," do
18 you see that?

19 A. Yes.

20 Q. And a click is literally just somebody using their mouse or
21 their finger to click on a result?

22 A. Yes.

23 Q. Okay. The third category listed is "scrolls." And that's
24 scrolling through the screen like you were discussing earlier?

25 A. I think so, but I'm not entirely sure.

1 Q. Is that one of the meanings of scrolls with respect to user
2 interaction data?

3 A. I'm not sure.

4 Q. And a "mouse hover," is that how long somebody hovers their
5 mouse over a result before clicking on it?

6 A. I think it can be how long their mouse hovers over some
7 element on the search page they may or may not click.

8 Q. Okay. And for each of these interactions and others,
9 Google tracks them; right?

10 A. Tracks them? Well, as -- if you go to Google and you start
11 interacting with the search page, then many of the things you do
12 are sent back to Google and logged. I'm not sure about the
13 scrolls one, and I'm not sure whether -- I'm not sure whether
14 they log mouse hovers. In principle, we could, I guess.

15 Q. Now, there's a second "interactions" arrow on slide 5, and
16 it's in the middle of the page, and it's going from the text box
17 back to the "search" cloud.

18 Do you see that?

19 A. Yes.

20 Q. And that arrow refers to Google feeding the logged user
21 data back into Google Search; right?

22 A. Maybe. I think that's a reasonable interpretation.

23 Q. Okay. We can go ahead and put slide 5 aside, and I would
24 like to talk to you briefly about some differences between
25 mobile and desktop user data. Okay?

1 A. Okay.

2 Q. Desktop queries and mobile queries, they're search intents
3 differ often; right?

4 A. I don't think the search intents for desktop and mobile
5 queries differ often. I'm not an expert on this, but my
6 impression is that the two query streams are fairly similar.
7 But again, it's not an area of expertise for me.

8 Q. Okay. Well, you're familiar with the term "search
9 intents"; correct?

10 A. Search intents? I suspect that that phrase is being used
11 in some specific way that I'm not sure I precisely know.
12 Usually, when someone does a search, we can talk about having --
13 their intent, in plain language what it is they're trying to
14 accomplish.

15 Is that what you mean?

16 Q. Yes. So when a user on their mobile phone or their
17 desktop, they enter a search, and they want to get a result
18 back; right?

19 A. They want something, uh-huh.

20 Q. And if a desktop user wants to get a certain type of result
21 back -- strike that. Let me ask a better question.

22 Sometimes people's search intents can differ, depending
23 where they're at, right, their location?

24 A. Their physical location?

25 Q. Yes.

1 A. I guess there could be cases like if you search for DMV and
2 you're in California, you probably want the California DMV, or
3 if you're in a different state, you want a different DMV.

4 Is that what you mean?

5 Q. Yes, or for example, if I'm in Rochester, New York, and I
6 look up Norton's Pub, then I probably want the local Norton's
7 Pub and not the one in Perth, Australia?

8 A. That would make sense.

9 Q. But that's one of the ways that location helps determine
10 somebody's search intent; right?

11 A. You mean how we might figure out what their intent is based
12 on their location?

13 Q. Yes, or figure out a user's search intent generally.

14 A. Right. So knowing a person's -- a person's location can
15 sometimes help understand what it is they're looking for.

16 Q. And sometimes mobile users and desktop users have different
17 intents?

18 A. I think that's probably true, yes. I think that's true.
19 I'm not sure I could just sort of list off a bunch of examples,
20 but I think that's probably true.

21 Q. Let's move on and talk about some of the basics of search.
22 The Court has previously heard an overview of this last fall.
23 So we're not going to go through everything again, but I want to
24 make sure we're on the same page. Okay?

25 A. Okay.

1 Q. So I would like to show you UPX204, which is a slide deck
2 titled "ranking for research." And this is another document
3 that's admitted.

4 You presented UPX204 to the research team at Google;
5 correct?

6 A. To -- so I presented this to certain researchers at Google
7 who were getting more involved with Search.

8 Q. Let's go to slide 47, which is Bates ending 4241, and it's
9 a chart that's titled "basic control flow."

10 Do you see that, Dr. Lehman?

11 A. Yes.

12 Q. So I want to start at the top right of the page where it
13 says "web documents."

14 You agree with me that there's a lot of documents on the
15 Internet?

16 A. Yep.

17 Q. And there's an arrow pointing down to "crawl and indexing,
18 prepare for fast retrieval and scoring"; right?

19 A. Yes.

20 Q. So basically, Google crawls the web to get information to
21 use in its search product?

22 A. Yes.

23 Q. And that information is arranged and categorized, which
24 makes it easier to search?

25 A. Yes.

1 Q. And the top left of slide 47 says "search query," and
2 that's the search query entered into Google by the user?

3 A. That's right.

4 Q. Then there's an arrow pointing down into "query rewriter."
5 Query rewriter interprets the query; right?

6 A. Yes. I think these things are a little bit dated. There's
7 also some flows from the search query on other paths. But yes,
8 query rewriter still exists, and that's its function.

9 Q. Okay. And interpreting a query generally or annotating a
10 query means adding additional information to it to help come up
11 with relevant search results? Would that be a fair
12 characterization?

13 A. I think that's fair.

14 Q. So that's stuff like adding synonyms to somebody's search?

15 A. Yes, that would be an example of an annotation in a query
16 rewriter.

17 Q. Now, the middle of slide 47 shows the query interpretation
18 process meets the crawling and indexing process, and it's at
19 that point in time where documents are ready to be searched;
20 right?

21 A. Yes. Again, there are some other paths now.

22 Q. Sure. And this is very, very high level. I understand
23 it's much, much more complex.

24 A. So I will stop repeating that if we just agree on it.
25 Okay.

1 Q. So once Google has -- the documents are ready to be
2 searched, they're retrieved, and they're scored; correct?

3 A. Yes. There's a fair amount of complexity there, and that's
4 probably a bit of a simplification, but maybe that's good enough
5 for now.

6 Q. And the documents, the scored documents are then ranked?

7 A. Well, let's see. So in ascorer, documents are given -- are
8 called ascorer scores. And then in the next step, which is now
9 highlighted in yellow, "super root," there are additional
10 adjustments to those scores. And then typically, search results
11 are ordered according to those scores in decreasing score order.
12 There are certain exceptions, but that's sort of the typical
13 case.

14 Q. There's also a step in the basic control flow
15 called "packaging"; right? Documents are packaged together?

16 A. Packaging?

17 Q. Like they're put together before they're sent on to the
18 user?

19 A. Maybe the packer? I think that -- if that's the component
20 you're referring to -- and I'm not super familiar with this, but
21 I think the job of the packer is to arrange things visually on
22 the page.

23 Q. And then search results are returned to the user after they
24 finish through these processes; right?

25 A. Yes, and then they're sent -- the sort of -- once the page

1 is sort of visually constructed, it would be sent to the user.

2 Q. Now, slide 47 of UPX204 does not show the logging of the
3 web search results; correct?

4 A. That's right.

5 Q. We can go ahead and put UPX204 aside. And I want to talk a
6 little bit about how Google evaluates its quality as a search
7 engine. Okay?

8 You would agree with me that high-quality search engines
9 should return high-quality search results?

10 A. Yep.

11 Q. And you would agree that there's many aspects of a
12 high-quality search engine?

13 A. Yes, there are many aspects of -- many aspects of a high-
14 quality search engine or high-quality search results.

15 Q. So relevancy is one of those important aspects of a high-
16 quality search engine; right?

17 A. I think when thinking about sort of the value of the search
18 result for a query, relevance is the most important
19 consideration.

20 Q. Because if I look up "DR," it matters if I'm looking for a
21 doctor or the Dominican Republic, as a fairly simplistic
22 example?

23 A. That would be sort of a case of an ambiguous query. Absent
24 any other information, either of those search results could be
25 relevant.

1 Q. So having more user data helps Google figure out what
2 results might be more relevant; right?

3 A. So having user data is useful to Google in identifying
4 relevant results for a search query.

5 Q. Okay. So like if I put in "DR vacation," it's more likely
6 I'm trying to go to the Dominican Republic on vacation instead
7 of looking for Dr. Vacation?

8 A. I think so.

9 Q. And we talked about localized search results. That's
10 something important to search quality; right?

11 A. When you say "localized search result," do you mean
12 something like showing the California DMV to somebody in
13 California?

14 Q. Yes, or where can I watch the Ohio State football game even
15 though I'm in the District of Columbia?

16 A. And the idea is that -- so in general, showing people
17 search results that are appropriate to their location for a
18 certain query is important, yeah.

19 Q. So let's go back to UPX204, which we just looked at, and
20 this time I'm going to show you slide 25. And this is a slide
21 titled "search quality has many aspects."

22 Do you see that?

23 A. Yep.

24 Q. And there's relevance and localization as two aspects for
25 search quality?

1 A. Yep.

2 THE COURT: What's the Bates number again?

3 MS. MURDOCK-PARK: This is UPX204 and the Bates ending
4 in 219.

5 THE COURT: Okay. Thank you.

6 BY MS. MURDOCK-PARK:

7 Q. There's other aspects of search quality listed on slide 25
8 that we haven't talked about, like privacy, for example?

9 A. Privacy, yes.

10 Q. And the bottom of slide 25 says "capturing everything in a
11 metric is tough"; correct?

12 A. Yes.

13 Q. Okay. So Google uses different yardsticks to measure how
14 well it's meeting its users' needs?

15 A. So currently, at least when I left Google, the primary
16 measure of search quality is a metric called IS4, information
17 satisfaction version 4, the metric, and maybe it's a higher
18 number now. That captures many of these aspects of search
19 quality, but it can't get quite all of them. And so sometimes
20 we use other methods to compliment IS4.

21 Q. Okay. So the IS metric, information satisfaction metric,
22 that's human rater-based evaluation; right?

23 A. That's right, yep.

24 Q. So it's basically paying people to evaluate search results?

25 A. Yes, though Google doesn't make those payments directly.

1 It works with an outside vendor.

2 Q. Okay. And most changes in ranking over the past few years
3 before you left Google, most changes in ranking tried to
4 maximize the IS score; right?

5 A. So yes, but I think something we always try to teach search
6 engineers is that making great search is not a game of
7 maximizing metrics. So metrics can be a useful guide towards
8 great search, but there are a lot of little games you can play
9 to maximize metrics that actually probably don't make the search
10 engine more useful to people.

11 So generally, yes, we hope to find changes that increase IS
12 but in ways that are actually beneficial to people.

13 Q. Now, you mentioned another metric, the live experiment, or
14 LE, metric; right?

15 A. That's not a single metric. It's an experimentation
16 process.

17 Q. Okay. And it's -- that's user-based evaluation basically;
18 right?

19 A. Yes.

20 Q. Okay. So it's a live traffic experiment that runs on a
21 percentage of users?

22 A. Typically, or -- so if we want to experiment with some sort
23 of change to the search engine, then we would by various means
24 expose a subset of Google users to that change, gather data that
25 gives us some insight into the nature of that change.

1 Q. Okay. So basically, turning on a new feature in realtime
2 and seeing how users seem to interact with it?

3 A. So a live experiment is sort of like turning it on,
4 although it can be a bit more complicated than that.

5 Q. The information satisfaction, or IS, metric and the LE
6 metric don't always agree; right?

7 A. That's correct. Sometimes they disagree.

8 Q. And that's because IS and LE give Google different
9 information about user behavior?

10 A. Well, the goal of the metrics is not to measure user
11 behavior. The goal of the metrics is to measure quality of
12 search results.

13 Q. Okay. And how users behave or interact with those results;
14 right?

15 A. No. The purpose of these metrics is to measure how good
16 our search results are and whether or not they're meeting
17 people's needs.

18 Q. Okay. So if we could go to UPX192, and this is another
19 document that is admitted. I'm going to ask you to look in your
20 binder at page 2, and we're not going to put that up on the
21 page, but just to confirm that you sent the e-mail at 1:32 p.m.
22 on page 2 of UPX192.

23 A. Okay. So this is UPX192.

24 Q. Yes, and the second page.

25 A. Second page. Okay. And you -- so it looks like this is an

1 e-mail thread -- I sent an e-mail, and there's a response or
2 something like that. Okay. Then you're saying on the second
3 page, you want to know the date and time of that particular
4 chunk in kind of the middle part of the second page?

5 Q. I'm just confirming that that -- you wrote the e-mail,
6 including the bullet points, that's on page 2; correct?

7 A. Yes.

8 Q. All right. And then there's some links that are on page 1
9 of UPX192; right?

10 A. Yes, that's correct, there are some links.

11 Q. So I would like to go to slide 4, page 4, which is the
12 first slide of the deck titled "unified click prediction."

13 Do you see that?

14 A. Yes.

15 Q. And this is a slide deck that you authored?

16 A. Yes.

17 Q. Let's go to slide 24. And it's titled "approximately 1
18 million IS ratings."

19 A. For slide 24, do you know the --

20 Q. The end is 787.

21 A. Okay.

22 Q. Okay? We're at slide 24?

23 A. Yes.

24 Q. Okay. And there's a really fuzzy picture on the right side
25 of the slide; correct?

1 A. That's right.

2 Q. You write in the second paragraph that IS ratings give only
3 a low-resolution picture of how people interact with search
4 results; correct?

5 A. Yes.

6 Q. And so you're conveying -- well, strike that.

7 In the third paragraph, you write, "Useful behavioral
8 patterns may appear in few training instances and thus cannot be
9 learned from IS ratings"; right?

10 A. Yep.

11 Q. So you're conveying that information satisfaction is an
12 incomplete evaluation metric by itself?

13 A. No. I think that's misinterpreted. What I'm talking about
14 here is to what extent can you learn about language and the
15 world from these two different data sources. It's not
16 commenting on the quality of the information satisfaction metric
17 as a tool for evaluating the quality of search results.

18 Q. Okay. But you -- by itself, you can't get a full picture
19 of search quality results; would that be fair?

20 A. By itself, with information satisfaction alone?

21 Q. Yes.

22 A. It's the best available single metric we have for measuring
23 the quality of search results, but it's not perfect.

24 Q. But IS can only give you a fuzzy picture; right? That's
25 why you put the fuzzy picture on the slide?

1 A. So the point here, I think -- maybe I could just look at
2 this for a second.

3 Q. Certainly. And I'm going to ask you about slide 25 next.
4 So that might help clarify things.

5 A. Yes, okay, I've had a chance to look at this.

6 Q. So the IS metric itself doesn't give you a full picture
7 standing alone; right?

8 A. So the IS metric is the best metric we have for measuring
9 the quality of search results, but in certain -- there are
10 aspects of the quality of search results that it doesn't
11 capture.

12 Q. But IS -- well, strike that.

13 Let's go to the next slide, slide 25, which is
14 titled "approximately 100 billion clicks."

15 Do you see that?

16 A. Yes.

17 Q. And the right side of slide 25 has a high-resolution
18 picture of the earth; correct?

19 A. Right.

20 Q. So the first sentence reads that those 100 billion
21 clicks "provide a vastly clearer picture of how people interact
22 with search results"; right?

23 A. Yeah.

24 Q. And the second paragraph reads, "A behavior pattern
25 apparent in just a few IS ratings may be reflected in hundreds

1 of thousands of clicks, allowing us to learn second and third
2 order effects"; right?

3 A. Yep.

4 Q. So second and third order effects means clicks help give a
5 more in depth understanding of user behavior?

6 A. Let's see. No, I don't think that's quite right. I think
7 what these slides are getting at is if you're trying to
8 understand how people use language or trying to understand facts
9 about the world -- and let me give an example, just sort of
10 trivial. When people search for something involving a cat,
11 we -- there may be search results that are appropriate to show
12 that don't contain the word cat but do contain the word kitten.

13 So now within -- if you try to learn that fact from our IS
14 rating data, that would be hard. There just isn't enough data.
15 There wouldn't be very many examples of queries containing cat.

16 On the other hand, if you got a very large volume of click
17 data, then you may have many, many examples of people who have
18 queries that contain the word cat and clicking on search results
19 for -- that contain the word kitten.

20 So these slides are talking about trying to understand
21 language and the world through use of these two data sources.
22 It's not talking about these two metrics or these two approaches
23 to measuring the quality of search results.

24 Q. So then you would agree that 100 billion clicks gives a lot
25 of information to help understand user behavior; right?

1 A. So I guess if we define clicks as sort of a record of user
2 behavior, then, you know, 100 billion records of user behavior,
3 yeah, it tells you a lot about user behavior.

4 Q. Are you saying that the machine learning systems learn
5 better from search logs, if I'm understanding?

6 A. That's a complicated question. It's evolved over time. So
7 Google has used a variety of machine learning systems in search.
8 Many of them -- I don't know how to generalize, actually.
9 They're all over the board. Some use IS data. Some use click
10 data. Some use both. Some use neither.

11 Q. Do IS and LE metrics tell you different things about search
12 quality?

13 A. I think they provide two different perspectives on search
14 quality, with IS as, I think, probably the stronger guide to
15 search quality, but I think live experiments can often provide
16 useful additional information.

17 Q. Okay. We can go ahead and take down UPX204.

18 Earlier, we discussed the "Google is magic" slide deck.

19 A. Right.

20 Q. And I would like to go to a more recent version of the deck
21 that you created in May 2020. If we could go to UPX219, and
22 UPX219 is admitted. This is a deck titled "logging and
23 ranking."

24 A. Yes.

25 Q. And you wrote the logging and ranking deck; right?

1 A. Yes. And I think as you've noted, parts of it are drawn
2 from that earlier presentation.

3 Q. If we could go to the speaker notes on page 3.

4 THE COURT: I'm sorry. This was prepared when? 2020?

5 BY MS. MURDOCK-PARK:

6 Q. Mr. Lehman, UPX219 was prepared in May 2020; is that right?

7 A. I don't know about May. I think the earlier one was 2017.
8 This was during COVID because it was sort of a cheer-up
9 presentation for the logging team, but beyond that, I don't
10 quite remember.

11 Q. Was it early in COVID or late in COVID?

12 A. I don't remember unfortunately. Probably earlier in COVID.

13 Q. If we could go to slide 3. I want to draw your attention
14 to the last sentence that you wrote from your speaker notes,
15 which is, "Effective logging for ranking is central to search,
16 to Google as a whole, and beyond."

17 Do you see that?

18 A. Yeah. So I'm trying to tell the logging team that their
19 job, which is not necessarily the world's most pleasant or
20 glamorous job, it matters, it's important.

21 Q. So it matters to the entire company; right?

22 A. Let's see. I don't know about the entire company, but it
23 certainly matters to the Search team and, I think, the Ads team.

24 Q. Okay. Well, you wrote that "effective logging is central
25 to Google as a whole"; right?

1 A. I wrote that. In context, I'm trying to cheer up this team
2 stuck at home.

3 Q. Okay. If we can go to slide 30, which is Bates ending
4 2426, and it's titled "logs to ranking to money."

5 Do you see that?

6 A. Okay.

7 Q. Now, the bullets on slide 30 explain why ranking needs user
8 data so much; right?

9 A. Okay. I have to reacquaint myself with this. Sorry. It's
10 been a long time.

11 Okay. I've looked through it, yeah.

12 Q. The second bullet of the slide, you write, "Not just one
13 ranking system learns from search logs. Learning from logs is
14 the main mechanism behind ranking."

15 Right?

16 A. That's what it says, yeah. Again, remember the audience;
17 remember the situation.

18 Q. And you also write at bullet point 4 that "web ranking is
19 only a part of search, but many search features use web results
20 to understand what a query is about and trigger accordingly."

21 A. Yes.

22 Q. Right? And so search features, I believe we've mentioned
23 those before, but that's stuff like knowledge cards or facts?

24 A. Yeah, there's a lot of search features, translation things,
25 weather boxes, all kinds of things.

1 Q. And you also write that other parts of Google benefit from
2 user data, right, like Ads, YouTube, Play?

3 A. Yes.

4 Q. And then I would like to turn you to your speaker notes, in
5 the middle of the speaker notes. Actually, about a third of the
6 way through, you write, "Not one system but a great many within
7 ranking are built on logs. This isn't just traditional systems
8 like the ones I showed you earlier, but also the most
9 cutting-edge machine learning systems."

10 Right?

11 A. Yes, it says that.

12 Q. So both older Google ranking systems and new machine
13 learning systems leverage user data; right?

14 A. This is as of, I guess, some time around 2020 that was
15 true. Subsequently, some -- an important machine learning
16 system did not use any user data.

17 Q. Now, according to your notes, as of this -- certainly, in
18 this slide deck, you say that cutting edge machine learning
19 systems rely on user data to function; right?

20 A. It says cutting-edge machine learning systems, yes. Again,
21 that's cutting edge as of this time. The field has moved very,
22 very fast.

23 Q. So at the very bottom, you write that you're not in
24 finance, "but grossly speaking, I think a huge amount of Google
25 business is tied to the use of logs in ranking"; right?

1 A. Yes.

2 Q. We can go ahead and set the document aside for right now.

3 You would agree with me that user data is used in multiple
4 parts of Google Search processes, right, multiple parts of their
5 ranking components?

6 A. Right. There are many ranking components that make use of
7 what we are calling user data or clicks.

8 Q. Okay. And other components outside of ranking also utilize
9 user data; right?

10 A. I know less about those things, but it's the other
11 components. So I think the Ads team uses user data. I don't
12 really know how. That's their world. And outside of ranking --
13 you're saying in Search outside of ranking, like feature
14 triggering or something like that?

15 Q. Multiple different aspects.

16 A. Probably. I mean, I guess without naming something
17 specific, I'm not sure if there are multiple aspects. But user
18 data is widely used within Search. So I wouldn't be surprised
19 if there's something outside of ranking.

20 Q. And is it used directly in some systems and indirectly in
21 others?

22 A. Within ranking?

23 Q. Within ranking or within search as a whole.

24 A. So I guess I can speak of within ranking. I'm not sure it
25 would be direct versus indirect use of user data. I guess I'm

1 not sure what the difference is.

2 Q. Well, direct as in it's used directly to train a certain
3 components, and indirectly like there's a second component that
4 trains off the first.

5 A. There are components in Search which make use of user data,
6 and then there are cases where we will take, say, data from that
7 first system and sort of reprocess it and use it in another way,
8 yeah.

9 MS. MURDOCK-PARK: Your Honor, I'm about to get to
10 another document. I believe I have another 15 minutes
11 approximately. I'm happy to push through, or we can take our
12 afternoon break.

13 THE COURT: Why don't we go ahead and push through
14 until we get to the end of the public session.

15 BY MS. MURDOCK-PARK:

16 Q. Mr. Lehman, I'd like to show you another slide deck. This
17 is UPX1115, and it's titled "search quality all hands 2018."

18 A. Yes.

19 Q. All hands means all search quality team members; right?

20 A. Yeah, all members of the search quality team were invited.

21 Q. Okay. And you were either at this training or invited to
22 this training?

23 A. It's not a training really. It's sort of a get-together to
24 discuss plans and --

25 Q. Meeting?

1 A. Sure.

2 Q. Okay. But you would have attended or had this deck
3 available to you as one of the search quality team members?

4 A. It would have been available to me, yes, and I probably was
5 there.

6 Q. Okay. If we could go to slide 42, which is Bates ending
7 1529, and it's a slide titled "result previews and language
8 understanding."

9 Do you see that?

10 A. Yes.

11 Q. Okay. And it's got three arrows chasing each other around
12 the words "virtuous cycle"; correct?

13 A. Yes.

14 Q. Okay. People at Google refer to the use of user data
15 within Google Search as a virtuous cycle; correct?

16 A. I don't know if that's generally true.

17 Q. You've heard the term "virtuous cycle" refer to the cycle
18 of user data that -- the dialogue that you used in the magic
19 presentation; right?

20 A. Virtuous cycle used in connection with, let's see, the
21 magic presentation?

22 Q. Strike that. Let's start with, you've heard the
23 term "virtuous cycle" used at Google before; correct?

24 A. I'm sure -- I've probably heard it used in multiple
25 contexts, yeah.

1 Q. Okay. And you've heard it referred to with respect to
2 Google Search; right?

3 A. Yeah, probably multiple contexts even in connection with
4 Google Search, yeah.

5 Q. So try and break down the self-reinforcing cycle that we've
6 got on UPX1115. At the upper right side of the screen, there's
7 a smily face with a phone. And it's above a green arrow;
8 correct?

9 A. Okay. So let's see. Yeah, it's near the green arrow.

10 Q. And it says "happy users informed user interactions";
11 right?

12 A. Yes.

13 Q. And then the green arrow goes down to the blue arrow, and
14 it says "better training data"; right?

15 A. Yes.

16 Q. Underneath the blue arrow, there's a bullet that
17 says "better models (ranking, language understanding)"?

18 A. Yes.

19 Q. Then the blue arrow feeds into an orange arrow that
20 says "better result previews, better results"; right?

21 A. Yes.

22 Q. Okay. So then orange arrow goes back to the emoji with the
23 happy users and informed user interactions; right?

24 A. Yes.

25 Q. So happy users means better data. Better data means better

1 systems and better results. And those better systems and better
2 results lead to happier users; right?

3 A. It's a little bit different. So the goal here -- to
4 explain the context, when you do a search, you get a bunch of
5 search results, and those are called result previews, because
6 you don't actually know what's on those web pages until you
7 click them and go read.

8 So I think what this slide is talking about is the
9 importance of making sure those result previews give people a
10 good, well, preview of what they're going to find on that web
11 page when they click through.

12 So a really bad thing would be we make a web page look like
13 it's relevant to your query when it's actually not and you go
14 click and you read for 30 seconds and are like wait, this
15 doesn't answer my question. We're wasting people's time a lot.

16 So better result previews has two benefits. One is happy
17 users. We're not sending people off on useless web pages. And
18 it has a second benefit, which is more informed user
19 interactions. If people know what they're clicking on, then the
20 clicks become more meaningful.

21 As an extreme, you could imagine if we had a search result
22 page and we just said here are results 1 through 10 and gave you
23 no information, your clicks would be random and meaningless.

24 So by improving result previews, clicks contain more
25 information. So it isn't the happy users that give better

1 training data. It's the second bullet point, more informed user
2 interactions.

3 Q. Okay. But the slide refers to happy users; right?

4 A. Yes.

5 Q. And it refers to better training data going to better
6 models; right?

7 A. Yes, and that better training data is not coming from the
8 happy users. The happy users is an important side benefit. The
9 better training data is coming from more informed user
10 interactions because people can anticipate what's going to be on
11 the web page that they're clicking.

12 Q. The informed user interactions that Google then logs;
13 right?

14 A. Yes. So when people interact with Google, we log those
15 interactions.

16 Q. Okay. And by the orange arrow, it says "better result
17 previews, better results"; right?

18 A. Yeah. I think the idea, if we're sort of interpreting
19 this, is a cyclic thing. I think the upper bullet point can
20 kind of be seen as a starting point for going around the circle,
21 and the lower bullet point is kind of something you get on the
22 far side, I think. I'm not sure. I didn't make this slide.

23 Q. Fair to say that Google needs users to sustain the cycle
24 shown in slide 42? It needs those user interactions?

25 A. Let's see. So I have to confess, in looking at this slide,

1 I can follow the reasoning. Better result previews give you
2 some happy users because people click on results that are
3 relevant to them, and it also gives you more informed user
4 interactions.

5 Okay. I'm sorry. I'm trying to work through the reasoning
6 here.

7 I guess the thing is I don't quite understand why it's a
8 cycle. Because yes, a result of this process is we will get
9 better search results, but I don't see how that leads to better
10 result previews. So I guess I don't really understand why it's
11 a cycle.

12 Q. Well, my question is, the cycle breaks if there's no users;
13 right?

14 A. Yeah, and I guess I'm saying that I don't understand why
15 it's a cycle in the first place.

16 Q. We can go ahead and put UPX1115 aside.

17 Dr. Lehman, Google has a public position about how much
18 user data it utilizes; right?

19 A. I don't know that -- I'm not sure. So this might be
20 getting into some sensitive matters. I can just go ahead, or I
21 don't know what to do.

22 Q. How about I show you a slide which is not redacted, and I
23 will ask you some questions on it.

24 A. Okay.

25 Q. If we could go to UPX204, and this time to slide 14. This

1 is Bates ending 208, and it's titled "sensitive topics."

2 Do you see that?

3 A. Yes.

4 Q. Okay. And underneath the picture -- that's Sundar Pichai,
5 Alphabet's CEO; right?

6 A. Yep.

7 Q. Underneath the picture of Mr. Pichai, you write, "Do not
8 discuss the use of clicks in search, except on a need to know
9 basis with people who understand not to talk about this topic
10 externally. Google has a public position. It is debatable.
11 But please don't craft your own."

12 So Google has a public position on clicks; right?

13 A. Right. So I guess I will go ahead and talk about this. I
14 don't know whether there's trade secret issues here. I don't
15 really know where the boundary lies.

16 Q. I mean, if you're concerned, you can answer just yes or no.
17 I don't want to tell you how to testify, obviously, but --

18 A. So let's see. So Google has a public position on the use
19 of clicks in search, yes.

20 MR. SMURZYNSKI: Your Honor, I think we're clearly
21 getting into an area where Dr. Lehman is concerned about
22 revealing trade secrets of Google, and I think that's why we
23 have a closed session, and I think it would be appropriate to
24 ask him further questions on this in the closed session.

25 THE COURT: Okay. Why don't we have counsel just run

1 through whatever additional questions she thinks she can elicit
2 testimony on, and then we will see where we are at that point.

3 MS. MURDOCK-PARK: Thank you, Your Honor.

4 BY MS. MURDOCK-PARK:

5 Q. By Google's public position on clicks, you're referring to
6 how much user data Google actually uses in its systems; right?

7 A. No.

8 Q. Okay. Google's public position is not entirely accurate
9 about clicks; right?

10 MR. SMURZYNSKI: Your Honor, I think -- there's a
11 reason there's a public position, and there's a reason --

12 THE COURT: I would like to just find out what the
13 public position is. Why don't we start there, because it's
14 public.

15 BY MS. MURDOCK-PARK:

16 Q. Dr. Lehman, can you please state what Google's public
17 position is on clicks?

18 A. So I'm not a communications person, but I think the
19 position that we've strived for is to try not to tell search
20 engine optimizers and people who would like to manipulate our
21 search results for their own benefit, to try not to give them a
22 really clear and unambiguous message that we use user data,
23 because they could potentially manipulate it to damage Google
24 and damage the quality of search results for everyone who uses
25 Google.

1 So we have said that we use user feedback data, clicks, in
2 evaluation of our search systems. And we try to avoid
3 confirming that we use user data in the ranking of search
4 results. So this has included things, I think, like saying
5 well, there are a lot of challenges with using clicks in ranking
6 of search results, not outright saying we don't do it but not
7 encouraging the belief that we do.

8 The issue here is that pretty much everybody knows Google
9 uses clicks in ranking. We're already being attacked. So there
10 are occasionally people who say like wait, why are you guys
11 trying to be secretive about this, because seriously, everybody
12 knows this.

13 And that's the thing, is please don't craft your own
14 position. And that's why it's debatable. And that's the debate
15 people have. They say why are you trying to obscure this issue
16 when it's totally not obscure, everyone knows it.

17 MS. MURDOCK-PARK: I think that ends the questions I
18 have in public session, Your Honor.

19 THE COURT: Terrific. It's a little after 3:15. We
20 will resume at 3:35.

21 Dr. Lehman, I will ask you to not discuss your testimony
22 with anyone during the break.

23 THE WITNESS: Okay.

24 THE COURT: Thank you, sir.

25 (Recess taken from 3:16 p.m. to 3:41 p.m.)

1 (The following occurred under [REDACTED])

2 THE COURT: All right. Continue with the examination.

3 BY MS. MURDOCK-PARK:

4 Q. Welcome back, Dr. Lehman.

5 THE COURT: Just to confirm, anybody in the courtroom
6 who is not affiliated with either of the parties -- or any of
7 the parties, I should say? Perfect. Thank you.

8 BY MS. MURDOCK-PARK:

9 Q. Dr. Lehman, I would like to talk to you now a little bit
10 more about Google's use of clicks in some of its ranking
11 systems. Okay?

12 A. Okay.

13 Q. Let's go to UPX213, which is a document titled "outline of
14 ranking newsletters." This will be to the back half of your
15 binder in the "closed" section.

16 A. Got it.

17 Q. You wrote UPX213 in 2017; right?

18 A. The only ambiguity about that is I think I did an initial
19 version of this and then did some revisions. It looks like at
20 least the last revision was around 2017, and possibly, I did
21 parts of it earlier in 2016.

22 Q. Okay. But if you look at the end of the first page, the
23 first underlined bullet -- actually, if we could pull out of
24 that. If we could go to the very bottom of the page, it shows
25 that there is a newsletter under "evaluation," and it's from

1 March 2017.

2 Right?

3 A. Yes.

4 Q. So it's at least some time March 2017 or later; right?

5 A. That would be the revision, yeah.

6 Q. Okay. And UPX213 is basically an encyclopedia of web
7 ranking; is that fair?

8 A. I think it's kind of an outline. It doesn't go into very
9 much detail itself. I just tried to put in a few brief
10 sentences. And I'm linking out to newsletters, which are each
11 themselves anywhere from a few pages to 20 or 30 pages long.

12 Q. Okay. And you write that this is an outline of ranking
13 newsletters which cover one technical topic each week. This
14 outline and the underlying articles may serve as a sort of
15 encyclopedia of web ranking; right?

16 A. Uh-huh.

17 Q. Okay. And you wrote some of the newsletters that are in
18 the outline?

19 A. I have written some newsletters. I don't know. They may
20 be after this. I'm not sure. I could have.

21 Q. Okay. So let's go to the middle of page 10, which ends
22 5723, and there's a heading titled "clicks."

23 Do you see that?

24 A. Yes.

25 Q. Okay. And you write, "Exploiting user feedback,

1 principally clicks, has been the major theme of ranking work for
2 the past decade."

3 A. Yes.

4 Q. So since at least 2007, using clicks had been an important
5 theme in ranking; right?

6 A. Yes, it had been an important theme.

7 Q. Okay. So earlier than 2007 even?

8 A. So this is referring to a particular approach to using
9 clicks. Oh, no, it isn't. Yeah, I think clicks were first used
10 before I joined Google in 2005.

11 Q. Okay. And you write next, "One can regard each as a
12 massive multiple-choice test. Each day we get to ask humanity a
13 billion questions of the form, which of these ten documents is
14 most relevant to your query? And then some person hopefully
15 answers with a click, which teaches us something about the
16 world."

17 Right?

18 A. Yes.

19 Q. And it teaches Google something about the world because
20 with each click users provide some data to answer the question
21 about the usefulness of a result?

22 A. Let's see. So a user -- so somebody has a query, right,
23 and we show ten results. And then maybe somebody clicks on a
24 search result in the simplest case. That doesn't mean it's a
25 great search result, but it's telling us something. They chose

1 to click on it.

2 Now, the question is, how does that help us learn about the
3 world? Because by itself, a click just tells us that this query
4 that we can't read provoked a click on this search result that
5 we can't read with computers or we couldn't read.

6 However, by looking for patterns in clicks, you can start
7 to learn things about language and the world. I think we talked
8 about the cat/kitten example earlier.

9 Q. And over time, Google has been able to get, as you write,
10 "extracting more general conclusions"; right?

11 A. That's right. So I can explain a little more about that if
12 you want. It's kind of an important theme.

13 Q. Yes, please.

14 A. So when people work with this kind of data and other types
15 of data sets unrelated to search, they talk about memorization
16 versus generalization.

17 So early use of click set Google was focused on
18 memorization. So each click says, okay, we've got a query, and
19 there's a search result, and it got a click. So it's very, very
20 specific. And, you know, there are trillions of possible
21 queries, close to a trillion possible documents, a huge number
22 of combinations.

23 So each piece of data you get is very, very specific, and
24 you need a certain number to sort of cover all the cases that
25 are going to arise in the future. It's just completely

1 impossible.

2 So that's memorization. But you can begin to look for
3 patterns in clicks. It could be that -- like you might notice,
4 for example, that if the query contains "06," people are pretty
5 happy clicking on search results that say "2006." And you might
6 sort of begin to realize oh, wait a minute, there's some kind of
7 connection here.

8 Or -- well, there's a lot of patterns like that. So by
9 finding patterns in the click data, you can begin to generalize,
10 and generalization systems extrapolate from situations where we
11 have data to situations where we don't. In effect, they try to
12 sort of -- you can think of it as filling the holes of Navboost
13 or reinforcing Navboost where there isn't much Navboost data or
14 just completely removing a need for Navboost perhaps. Navboost
15 is a memorization system. Sorry about that.

16 So the work that was done over this sort of decade was on
17 generalization systems, trying to extract the patterns from
18 clicks. And in the extreme, which is more recently, if you find
19 enough patterns in clicks or in language, then the hope is
20 eventually you could start to get computers to read themselves.

21 Q. You mentioned Navboost. If we could actually pull out of
22 this click section, and there's some sort of subheadings, and
23 Navboost is the second one down on the page; right?

24 A. Yes, that's right.

25 Q. Okay. So you said that Navboost, it was a memorization

1 system?

2 A. That's right, yeah.

3 Q. And you write that "Navboost is the original click
4 exploitation system and still the most potent"; right?

5 A. Yeah.

6 Q. Navboost trains on about [REDACTED] of user data; right?

7 A. That is my understanding. Now, the word "trains" here
8 might be a little misleading. Navboost is not a machine
9 learning system. It's just a big table. It says for this
10 document -- sorry, for this search query, this document got two
11 clicks. For this query, this document got three clicks, this --
12 and so on. And it's aggregated, and there's a little bit of
13 extra data. But you can think of it as just a giant table.

14 Q. And Navboost is used in other systems or other ranking
15 components in Google; right?

16 A. Yes. So some -- what's used in other ranking components?
17 So Navboost is both a sort of set of data and itself is a
18 ranking component that uses that data to adjust the scores of
19 search results when it's available.

20 Navboost can be used indirectly, say, in triggering certain
21 features within search. So yeah, it's used in a variety of
22 places.

23 Q. And we can go ahead and put UPX213 aside.

24 You're familiar with the component or the system Glue;
25 right?

1 A. I'm somewhat familiar with Glue.

2 Q. Can you explain at a high level what Glue is for the Court?

3 A. Okay. So Glue, I think, is well understood as sort of
4 relative to Navboost. So Navboost is a system --

5 THE COURT: Did you say glue or blue?

6 THE WITNESS: Glue, g-l-u-e, like the sticky stuff.

7 THE COURT: Okay.

8 THE WITNESS: So Navboost records clicks on search
9 results for queries, but there are other types of interactions
10 with a search page, and there are other things on a search page
11 besides just web search results. There can be all kinds of
12 little boxes with data and images and all that kind of stuff.

13 And Glue attempts to record all those other interaction
14 types on all those other elements of the search page for
15 different queries. That's my understanding. I'm not an expert.

16 Q. So Glue is also trained on user data, then; right?

17 A. So again, it's not a machine learning model. It's a giant
18 table. But yes, it's learning from user data. It's a super set
19 of Navboost.

20 Q. Okay. If we could go to UPX5, which is a slide deck
21 titled "Navboost/Glue: Life of a user interaction." I would
22 like to turn your attention to slide 6. The slide is
23 titled "Glue data in prod."

24 Do you see that?

25 A. So this is 808, I guess.

1 Q. Yes.

2 A. Okay. And it's titled "Glue data in prod," yes, uh-huh.

3 Q. Could you explain, again at a high level, what slide 6
4 represents with respect to Glue and other systems and ranking?

5 A. Okay. So I think what this is showing -- they're calling a
6 Glue model. It's not a machine learning model. It's a big
7 table of data. And this data would look -- a typical entry in
8 this Glue data would be something like a person did this search
9 query, and they looked for a long time at the sort of our
10 attempt to answer the question immediately with a paragraph on
11 the top of search results, or a person did this query, and they
12 scrolled through the list of newspaper articles.

13 So that's what's in the white box there, "Glue model."

14 The [REDACTED] stuff, we can get into it if you want.

15 And I think what this is showing is various systems where
16 that data is used. I think that's it.

17 Q. Okay. We can go ahead and set UPX5 aside. And I want to
18 ask you briefly about one more system at Google, which is QBST.
19 And "QBST" stands for query-based salient terms?

20 A. That's right.

21 Q. Okay. Can you explain, again briefly, what QBST does?

22 A. Let's see. So it is a system built upon layers and layers
23 of technology developed over many years by many engineering
24 teams. It's kind of a culmination of many previous efforts.
25 Generally, what you see when you use QBST is you stick in a

1 search query, and you get out a list of words and pairs of
2 words, which are usually called bigrams in the context of
3 search, and those words and pairs of words that come out of QBST
4 when you stick in a query are words that we think should appear
5 prominently on web pages that are relevant to that query.

6 I don't know if that's clear.

7 Q. Would terms like "1600 Pennsylvania Avenue" and "White
8 House," if you searched for one, you might get results for the
9 other, for example?

10 A. Yes.

11 Q. And QBST helps identify relevant documents to respond to
12 queries; right?

13 A. Yes.

14 Q. And QBST is trained on [REDACTED]
15 [REDACTED] user data; is that right?

16 A. I -- so it uses user data. I don't know -- and pretty much
17 every signal we have uses user data, the contents of documents,
18 links between documents, and lots and lots of algorithmic
19 processing. But I don't know the time span for the user data
20 that it uses.

21 Q. Navboost and QBST are the two most effective ranking
22 systems in figuring out language nuances; right?

23 A. No.

24 Q. Sitting here today, you --

25 A. Not even close.

1 Q. Okay. And do you recall being deposed, Mr. Lehman?

2 A. Yep.

3 Q. Okay. And you were deposed over two days?

4 A. Yep.

5 MS. MURDOCK-PARK: Okay. If I may approach, Your
6 Honor.

7 THE COURT: Okay.

8 BY MS. MURDOCK-PARK:

9 Q. I'm going to hand you a deposition binder and point you to
10 the April 22nd, 2022, deposition that you took and specifically
11 at page 398 -- I'm sorry, 397 on to 398.

12 A. Okay.

13 Q. And you were under oath during your deposition; right,
14 Mr. Lehman?

15 A. Yep.

16 Q. Dr. Lehman?

17 A. Yep.

18 Q. And at your deposition, you were asked, starting at line 23
19 of page 397:

20 "Question: So it is your position that Navboost and QBST
21 are both capable of memorizing truly long-tail information on
22 relevance and user preference?"

23 "Answer: Yes. Navboost and QBST are the -- the two
24 systems that we would use for -- rather, Navboost and QBST are
25 the two systems in our ranking system that -- most effective in

1 memorizing little facts about the world, little nuances of
2 language."

3 MR. SMURZYNSKI: Your Honor, I object to the
4 characterization of this as impeachment. This is not
5 inconsistent in any way.

6 THE COURT: Well, I --

7 MS. MURDOCK-PARK: If I may, Your Honor, I asked
8 Mr. Lehman whether --

9 THE COURT: There may be some nuance to it, but I
10 think it's appropriate impeachment. It's something you can
11 clean up on your examination.

12 MR. SMURZYNSKI: Thank you, Your Honor.

13 THE COURT: Go ahead.

14 BY MS. MURDOCK-PARK:

15 Q. And you were asked that question and gave that answer at
16 your deposition?

17 A. That's right.

18 Q. We can go ahead and put your deposition aside.

19 A. I can clarify the difference, if you would like.

20 Q. Your attorney can clarify that when --

21 A. Okay.

22 Q. -- he asks you questions.

23 But you've referred to long-tail information; right?
24 Strike that.

25 You've heard the term "long-tail information"?

1 A. Long tail information? So long tail is a term that I'm
2 very familiar with. I've heard of long-tailed queries. I'm
3 less clear about what we mean by long-tail information.

4 Q. Could you explain for the Court what a long-tail query is?

5 A. So it's kind of an informal term, but it sort of means a
6 rare query. I think the reason people say long tail instead of
7 rare is to refer to sort of a particular characteristic of the
8 query stream, which is that -- it's a little hard to describe
9 briefly, but -- okay, so there's some queries that are super
10 common, a small number of those. There are more queries that
11 are fairly common. And there are still more queries that are a
12 little bit less common and so on.

13 You can imagine tiers until at some point you get down
14 to -- well, it doesn't end, but later, you get to tiers of
15 queries that are individually extremely rare, but there's just
16 enormous numbers of them, like eight-word queries or something
17 like that.

18 And the characteristic of the query stream is that all of
19 these tiers are kind of equally probable, roughly, very roughly.
20 So there are these masses of extremely -- of queries that are
21 extremely rare individually, but collectively, they make up a
22 significant part of the query stream.

23 So that's sort of the long-tail phenomena. So anyway,
24 that's why they call them long-tail queries as opposed to just
25 rare queries.

1 Q. And the hype queries, those are the most common queries
2 that you were referring to?

3 A. Yes.

4 Q. And then there's torso queries? That's sort of the middle
5 of the road, not too hot, not too cold?

6 A. Yeah. These terms are all informal.

7 Q. I would like to turn back to one of the topics we discussed
8 in the public session, which is evaluation. We talked about the
9 human rater evaluation metric called IS, and IS is measured by
10 points; correct?

11 A. Points, yes, that's right.

12 Q. So one IS point is considered a lot in terms of an
13 improvement to search quality?

14 A. So the IS scale ranges from, I think, typically around 0 to
15 100. So if you're just starting and building a search engine
16 and you're kind of starting at zero, a gain of one is nothing.
17 For Google after it's been optimized for a quarter century, a
18 one-point IS gain would be like a good yearly goal.

19 THE COURT: Would be a what?

20 THE WITNESS: A good yearly goal.

21 BY MS. MURDOCK-PARK:

22 Q. Moving the IS score by half a point a year would be a big
23 jump in quality; would that be fair?

24 A. Moving the IS score by half a point?

25 Q. By half a point, yes.

1 A. So I think we would be a little disappointed in ourselves
2 at that rate. But I haven't tracked it super closely, but I
3 think that would be kind of on the lower end of what we would be
4 targeting. And different people may have different views.

5 Q. You had mentioned IS4 earlier, I believe.

6 THE COURT: I'm sorry. I don't quite understand. Is
7 the IS score, are you talking about a system-wide IS score, or
8 when you're talking about movement by -- in reference to what?

9 THE WITNESS: Right, right, yeah, and I think it's
10 confusing, because it's in reference to all of those.

11 So in a typical evaluation, there will be many, many
12 queries that are being evaluated at once, so a long list of
13 queries. For each of those queries, there's going to be a list
14 of search results, and in a typical evaluation, we only take the
15 top 5.

16 So then we send that query and those search results off to
17 raters, and they assign IS scores to each of those five search
18 results. So at that level, a single result for a single query
19 within this larger experiment.

20 And then we begin to aggregate, so up to the query level.
21 So for a given query, we take a kind of average of the IS scores
22 for those top 5 results, but it's weighted so the top results
23 count more. And there's a few other little nuances.

24 And then we aggregate up to the level of the whole
25 experiment. So we average across all the queries in the

1 experiment. So it can be at those three different levels.

2 THE COURT: Thank you. And you said the scale is
3 from?

4 THE WITNESS: I think pretty much 0 to 100.

5 THE COURT: Okay. Thank you. That's a very helpful
6 explanation.

7 BY MS. MURDOCK-PARK:

8 Q. You mentioned the top 5 positions in the IS results, right,
9 the results page?

10 A. Yes.

11 Q. Is that known as IS4 at 5, a metric that measures those top
12 5 results?

13 A. Yeah, the "at" usually refers to the number of results that
14 are being evaluated.

15 Q. And the first five positions of search engine results can
16 include web results or search features; correct?

17 A. So there are usually two different types of evaluation. So
18 people will say things like IS4 versus IS4 web. And so IS4 web
19 would just contain web search results; whereas, IS4 would
20 include both web results and features.

21 Q. Okay. So IS4 at 5, what would IS4 at 5 then represent in
22 terms of results?

23 A. So it would be an evaluation that measures the top 5
24 results where those results may be both web pages and other
25 search features.

1 Q. Google previously used the precision metric to measure
2 search quality; right?

3 A. Yes.

4 Q. I would like to pull up UPX974, which is a March 31st,
5 2010, e-mail from you. And this is in evidence.

6 Do you have UPX974 before you, Dr. Lehman?

7 A. Yes.

8 Q. Okay. And the subject of the e-mail is "precision code
9 orange."

10 Is a code orange a focused effort to address a pressing
11 issue?

12 A. It's kind of an informal notion, but yeah, typically, it
13 would be.

14 Q. Let's start with the e-mail from Amit Singhal on page 5.
15 Dr. Singhal is the former head of Google Search?

16 A. He was the former head of the Google ranking team, and I
17 think for a time he was also head of Search, yes.

18 Q. He used to be your boss?

19 A. For a short time, he was my boss. Other times, there were
20 one or two people between us.

21 Q. And Dr. Singhal wrote in his e-mail, "We have a serious
22 competitive threat from Bing in ranking. We need to act fast
23 and act decisively. I would like you to drop everything you are
24 doing starting today and debug the new 20K set for losses to
25 Bing (possibly ideal). Eric L. will coordinate."

1 Right?

2 A. Yes.

3 Q. So Mr. Singhal asked the Search team to drop everything to
4 address Bing's competitive threat?

5 A. He asked, it looks like, the people listed here.

6 Q. So one of the people that he listed there on his e-mail is
7 you; correct?

8 A. Yes.

9 Q. And he asked, "I would like you to drop everything you are
10 doing starting today and debug the new 20K set for losses to
11 Bing"; right?

12 A. Right, right. So the distinction I'm drawing is I thought
13 you asked like whether he asked the whole Search team to do it,
14 and it seems like it's this group of eight or ten people here.

15 Q. Okay. The eight or ten people who are in the Search team;
16 right?

17 A. Yeah.

18 Q. And you were asked to coordinate the effort; correct?

19 A. Yes.

20 Q. I would like to go to page 2, and there's an e-mail from
21 you at 11:19 a.m. It's at the bottom of the page. And you were
22 responding in the e-mail chain, and you write, "We have one week
23 to come up with a bunch of ways to improve search quality."

24 Right?

25 A. Yes.

1 Q. So in response to Bing's threat, you had to come up with
2 many different ways to improve search quality?

3 A. That was the strategy that we tried, yes.

4 Q. And before we move on from evaluation, I just have one more
5 question about the IS4 metric.

6 Does the IS4 metric measure ads?

7 A. No.

8 Q. Okay. So now I would like to move on, and we can go ahead
9 and put the document aside. I would like to talk a little bit
10 about Google Search compared to Bing Search. Okay? So when you
11 worked at Google, Google sometimes compared its search engine
12 quality to Bing's search engine quality; right?

13 A. Let's see. When I worked on Google Search, we sometimes
14 compared, yes.

15 Q. And for Google to make a meaningful comparison to Bing, it
16 needed to evaluate a fairly large number of queries; right?

17 A. Well, the larger number of queries, the smaller the error
18 bars get. I don't know what number was required.

19 Q. But there needed to be a statistically significant
20 comparison?

21 A. In order for the comparison to be meaningful, it would need
22 to be statistically significant.

23 Q. Okay. And you considered a large-scale comparison of Bing
24 and Google Search results to be difficult; right?

25 A. It depends on how the comparison is being done.

1 Q. In terms of comparing, for example, search features, Bing
2 and Google have different search features and different amounts
3 of search features; right?

4 A. I think so.

5 Q. Okay. And making a comparison between Google Search and
6 Bing Search requires a lot of engineering time and effort; would
7 that be fair?

8 A. Making comparison between Google Search and Bing Search, I
9 think it depends on how good the comparison is done. Yeah,
10 there are lots of ways to do it, some that are very labor
11 intensive and probably some that are not.

12 Q. Okay. For example, scraping, you've heard the term
13 scraping?

14 A. Yes.

15 Q. That would be a difficult, time-consuming effort, to scrape
16 Bing's data?

17 A. So scraping, as I understand the term, means -- in
18 connection with Bing would mean sending queries to Bing, getting
19 back -- and then getting back the search results. And that
20 process alone is not super difficult, although there are some
21 peculiar challenges on the block scrapers.

22 Once you've got the search results from Bing, trying to
23 identify well, what are the search results within this big mass
24 of HTML that describes the search results page, I think that's a
25 challenging thing.

1 Q. Okay. And it's hard to make a true apples-to-apples
2 comparison with Google Search and Bing Search; fair?

3 A. This isn't something where I have expertise. I have not
4 worked on evaluation.

5 I think for web results, it's probably a lot easier than
6 once you start to include all kinds of features, but this is
7 really not my area. I'm not sure.

8 Q. I'd like to show you UPX268. And I want to start at slide
9 2 or page 2, which begins a slide deck titled "Google Search
10 competitive fact pack - initial insights."

11 Do you see that?

12 A. Let's see. Did you say 268?

13 Q. 268, yes.

14 A. And then?

15 Q. Starting on page 2.

16 A. Ah, yeah, right, uh-huh.

17 Q. And the competitive fact pack is dated August 2020?

18 A. Yep.

19 Q. And the slide deck was shared with the Search quality team?

20 A. I don't know. I don't see any indication of that.

21 Q. Okay. Do you recall receiving the Google Search
22 competitive fact pack?

23 A. No.

24 Q. So I would like to show you a document that might refresh
25 your memory.

1 A. Okay.

2 MS. MURDOCK-PARK: If I may approach, Your Honor.

3 THE COURT: Okay.

4 MS. MURDOCK-PARK: This is, for the record, document
5 GOOG-DOJ-20246855.

6 BY MS. MURDOCK-PARK:

7 Q. And this is a Google calendar invitation, and it's
8 showing -- this was sent to your -- one of the people on the
9 e-mail is Pandu Nayak.

10 He was your boss; correct?

11 A. Yes.

12 Q. Okay. And it says -- subject is "competitive fact pack
13 share-out with Search quality team"; right?

14 A. Yeah.

15 Q. So that would be you; right? You're a part of the Search
16 quality team?

17 A. I'm one of a thousand people on the Search quality team.

18 Q. So we can go ahead and put that document aside.

19 I would like to go to slide 3 of UPX268.

20 MR. SMURZYNSKI: Your Honor, there's been no
21 foundation laid for use of this document with this witness. In
22 fact, the witness said he was unaware of having seen it before.
23 And if that refreshed his recollection, I didn't hear him say
24 so.

25 THE COURT: Is this already in evidence?

1 MS. MURDOCK-PARK: UPX268 is in evidence, Your Honor.

2 THE COURT: So since it's in evidence, she can ask him
3 questions about it. If he doesn't the answers, he doesn't know
4 the answers.

5 MR. SMURZYNSKI: Thank you, Your Honor.

6 BY MS. MURDOCK-PARK:

7 Q. Dr. Lehman, do you see the background of the competitive
8 fact pack at slide 3?

9 And the first bullet reads, "This work was primarily
10 conducted to assemble a fact pack on how Google compares to
11 other search providers across search metrics."

12 Do you see that?

13 A. Yep.

14 Q. Okay. And how Google compares with other search providers
15 is something that you would be interested in as a part of the
16 Search quality team?

17 A. Maybe a little, probably not very much.

18 Q. Okay. If we could go to slide 25, please. And slide 25 is
19 Bates ending 132, and it's a slide titled "features plus
20 ranking: Compared to Bing, Google is stronger in ranking while
21 Bing is stronger in feature diversity and interactivity."

22 Right?

23 A. Yes.

24 Q. And slide 25 shows some results for Google's search quality
25 compared to other search providers, including Bing, Yahoo!, DDG,

1 Ecosia, and Qwant?

2 A. So I see the headings there, Google versus Bing and Google
3 versus Yahoo!, and so on, yes.

4 Q. The bottom of the slide reads, "On desktop, Bing: Google
5 stronger in ranking, authoritativeness, local content, while
6 Bing is stronger in feature diversity and interactive results."

7 Right?

8 A. Yes.

9 Q. And if we could go to slide 26, which is titled "search
10 quality (features plus ranking): On mobile, Google leads all
11 search engines. On desktop, Google is comparable to Bing but
12 leads all others."

13 Right?

14 A. Yes.

15 Q. And the first column of slide 26 is a comparison of Google
16 and Bing; right?

17 A. I see Google and Bing there. I don't quite understand the
18 chart yet.

19 Q. Okay. If you look on the right-hand side, there's desktop
20 which is in red and mobile which is in blue; right?

21 A. Yes, I see that. Right, okay. I think I understand the
22 chart now.

23 Q. Okay. And there's a scale on the left side; correct?

24 A. Yes.

25 Q. And the box under -- or over "Google versus Bing," the

1 first box says [REDACTED].

2 Do you see that?

3 A. Yes.

4 Q. Looking at the scale on the left side of the screen, that
5 would be between Google slightly worse and about the same;
6 right?

7 A. Yeah. And it looks like it's much closer to about the
8 same; right? Yeah, I think so.

9 Q. Okay. And because there's no red desktop line on Google
10 versus Bing, the [REDACTED], that represents the desktop?

11 A. I don't know.

12 Q. The mobile, the blue is [REDACTED]; right?

13 A. Yes.

14 Q. Okay. And Google has a larger share of queries on mobile
15 than it does on desktop; right?

16 A. You mean within Google, we get more queries from mobile
17 than desktop?

18 Q. Between Google and Bing, Google gets more queries on
19 mobile; right?

20 A. So Google gets more mobile queries than Bing?

21 Q. Than Bing gets mobile queries.

22 A. I don't know. My impression is Google gets a lot more
23 traffic than Bing in general, but I haven't seen these numbers
24 in detail.

25 Q. Okay. We can go ahead and set the document aside.

1 Dr. Lehman, I want to turn very briefly back to some of the
2 user-based evaluation experiments that occurred when you were at
3 Google.

4 During your time at Google, do you recall any experiments
5 that involved Google limiting its use of search logs to be
6 equivalent to those of a competitor?

7 A. During my time at Google, I think I may have been sort of
8 vaguely aware that they were ongoing.

9 Q. Okay. So your testimony today is that you do recall
10 experiments that involved Google limiting its use of search logs
11 to be equivalent to those of a competitor when you were at
12 Google?

13 A. It's hard to remember exactly when I learned about
14 something now that happened in the past.

15 I think I was aware that there were experiments involving
16 reducing the amount of user -- sort of -- there are experiments
17 with creating a variant of Google Search that used less user
18 data. I think I was sort of vaguely aware that was happening.

19 Q. At the time of your deposition in April 2022, were you
20 aware of such an experiment?

21 A. I just don't remember when I became aware of these things.
22 I'm sorry.

23 Q. Okay. I would like to turn you, if I can perhaps refresh
24 your memory --

25 A. Okay.

1 Q. I would like to turn you back to your deposition
2 transcript, and this time we're going to look at the first day,
3 which is April 21, 2022.

4 A. Okay.

5 Q. And I'm going to draw your attention to page 183 and
6 specifically lines -- I'm sorry, 187, specifically lines 8
7 through 13.

8 A. Sorry. So 187 --

9 Q. Lines 8 through 13, and see if that refreshes your memory
10 of your awareness of experiments in April 2022.

11 A. Oh, 187. Sorry. I was looking at 186. Okay.

12 Oh, at the time I guess I was not aware of such
13 experiments. Okay.

14 Q. Okay. So we can go ahead and set your deposition aside
15 again.

16 Dr. Lehman, you're aware that Google retained Professor
17 Edward Fox as an expert in this case?

18 A. Yes.

19 Q. Did you speak with Professor Fox?

20 A. Yes.

21 Q. How many times?

22 A. It was either once or twice. I think once, but I'm not
23 sure.

24 Q. Do you recall when?

25 A. No.

1 Q. Did you take notes of your meeting with Professor Fox?

2 A. No.

3 Q. Did anyone suggest that you not take notes?

4 A. I don't remember.

5 Q. What did you and Professor Fox discuss?

6 A. So I think he wanted to know the amount of user data that
7 was used in some of the -- some of the machine learning models
8 that are used in ranking, and then I think he had just some sort
9 of basic getting oriented kinds of questions.

10 That's all I can remember right now.

11 Q. Did you talk about six ranking components in particular?

12 A. I don't -- we talked about several. I don't know if it was
13 six exactly. I don't remember that. I don't remember that
14 being like six something or other.

15 Q. Okay. Do you recall which ranking components you discussed
16 with Professor Fox?

17 A. Let's see. I think probably rank brain -- so our main
18 models, rank brain, deep rank, rank imbed. I think those, but
19 yeah, it's been a while. I don't remember.

20 Q. Do you recall if you discussed Navboost, QBST, or term
21 weighting with Professor Fox?

22 A. I don't remember one way or the other.

23 Q. Have you read Professor Fox's report?

24 A. No.

25 Q. Are you aware that Google engaged Professor Fox to conduct

1 an experiment where Google limited its use of search logs to be
2 equivalent to the share of a competitor and then measured those
3 results?

4 A. Would you say the first part of the question again?

5 Q. Sure. Are you aware that Google engaged Professor Fox to
6 conduct an experiment where Google limited its use of search
7 logs to be equivalent to the share of a competitor and then
8 measured those results?

9 A. I don't think I was aware of that outside the context of
10 this trial or sort of in connection with my regular work. I
11 don't know if that's --

12 Q. Going back to when you met with Professor Fox, do you
13 recall if you met with him before or after his experiment?

14 A. I don't think I knew about the experiment. So I don't know
15 if it was before or after.

16 Q. Did you tell Professor Fox that you had never done a data
17 reduction experiment yourself?

18 A. I don't recall telling him that. I just don't recall our
19 conversation that clearly.

20 MS. MURDOCK-PARK: Okay. I pass the witness, Your
21 Honor.

22 MR. CAVANAUGH: We have no questions for this witness.

23 THE COURT: Okay.

24 THE COURT: Mr. Smurzynski. I hope I'm pronouncing
25 that correctly?

1 MR. SMURZYNSKI: Yes, yes. It's intimidating, but by
2 the end of the trial, hopefully, Your Honor, it will flow.

3 THE COURT: My last name is Mehta. It's not, you
4 know --

5 MR. SMURZYNSKI: Could we just get organized for a
6 second, Your Honor? And then we will be ready to go.

7 THE COURT: Sure.

8 CROSS-EXAMINATION

9 BY MR. SMURZYNSKI:

10 Q. Good afternoon, Dr. Lehman. Do you recall being asked some
11 questions about a 2010 code orange with Dr. Singhal?

12 A. Yes.

13 Q. What was the role of a code orange within the Google Search
14 organization?

15 A. The role of a code orange within --

16 Q. Yes, yes.

17 A. I don't know how many there have been, but my understanding
18 of its role is that it would be sort of a concentrated
19 short-term effort to address some kind of a problem.

20 Q. And did a code orange change the amount of work that people
21 in the search organization were doing?

22 A. I don't recall this code orange changing the amount of work
23 that I was doing. It changed the nature of the work I was
24 doing.

25 Q. And what do you mean by that?

1 A. So -- okay. This was quite a long time ago. And so we --
2 we didn't have such well-established practices for just even
3 basic things. Like, how do you continuously improve a search
4 engine over years and years? We hadn't really worked out
5 exactly how to do that.

6 So one thing that we had been doing which seemed sensible
7 at the time was you'd look for sort of a group of queries that
8 seemed to have something in common, and then you would go away
9 for a while and try to devise some sort of algorithm to fix that
10 group of queries. And you would do some launch analysis to see
11 if you actually fixed them, and then you change the search
12 engine to introduce that fix. And then you just repeat that.

13 And the hope was that if you repeated that many times,
14 you're fixing that batch of queries, that batch of queries, that
15 batch of queries, then the quality of the search engine will
16 improve over time.

17 In some sense, that's true, but I think around the time of
18 that code orange, we realized that that approach wasn't working.
19 And the reason it wasn't working in retrospect -- I don't quite
20 remember when we got this all figured out -- was that it didn't
21 attend enough to the impact of each individual change. So it's
22 great if you can fix, I don't know, 0.1 percent of the queries
23 or make them better. But it's a whole lot better if you can
24 make 10 percent of the queries a little bit better.

25 So I think sort of prior to this code orange, we probably

1 had been using a little bit the wrong play book of trying to fix
2 many, many little problems. And we hoped that that would add up
3 to something significant, but it really wasn't. And I think we
4 sort of became aware of that around the time of this code
5 orange.

6 And we tried to shift practice. You can see us setting up
7 this process where we're going to try and look at lots and lots
8 of queries and find really big patterns of failure and go after
9 those and hopefully that will allow us to improve our search
10 engine faster.

11 I think in the discussion of that, there's a lot of we will
12 try this new strategy, see how it works, see if we learn
13 something.

14 So that's what was going on at that time. It's not like we
15 went from lasing about to working hard. Rather, it was we tried
16 to shift from one strategy to another. And I think what we
17 learned even later is that that strategy we shifted to, it was
18 probably better but still not the most effective way.

19 There isn't a play book for how to build a search engine.
20 We're writing it ourselves.

21 Q. Stepping back more globally, what motivated you when you
22 worked at Google Search?

23 A. I think a few things. I walked out of a Ph.D. in sort of a
24 pretty theoretical area. And this was an area where, after I
25 learned a bunch of new stuff, I felt like if I worked hard at my

1 job, I could do something that could at least in a small way
2 benefit a huge number of people around the world. And I found
3 that inspiring.

4 Another thing that, as years went on, I found increasingly
5 inspiring was that Google was, as a company, totally committed
6 to search, and search was this very, very deep problem. So it
7 was a place where you just burrow deeper and deeper.

8 And the basic problem at the core of search and determining
9 whether or not a web page is relevant to a search query, to
10 really, really do that, you have to get computers to read.

11 So that's how deep the problem is. It goes all the way to
12 AI. And particularly in more recent years, the connection
13 between search and AI became sort of increasingly clear. And
14 so, you know, burrowing very, very deep into this problem area,
15 it started to seem like wow, we're going after one of the most
16 fundamental problems there is, making computers think, while at
17 the same time doing something that I hope is beneficial to huge
18 numbers of people around the world.

19 That's all me.

20 Q. Did the presence of a competitor like Bing influence how
21 hard or how less hard you worked while you were at Google?

22 A. I think I put in as full days as I could manage while also
23 being a parent, and eventually, I couldn't manage it anymore,
24 and I had to just stop and be a parent.

25 Q. And I appreciate that answer.

1 Was the fact that at any particular time Bing was deemed
2 more competitive or less competitive, did that influence how
3 hard you worked at Google?

4 A. I think it changed the focus of my work, but it isn't like
5 I was putting in hugely different hours. I've still got to pick
6 up my kids from the day care.

7 Q. All right. Let's look at UPX004, which should be in your
8 binder. It should be the first document.

9 THE COURT: Counsel, is this something we can do in
10 open session?

11 MR. SMURZYNSKI: I think I may be asking questions in
12 more depth than the government perhaps did, and even though they
13 used a particular page from the document that is not
14 confidential -- I appreciate, Your Honor, and I'm desirous of
15 trying to accommodate as much of this trial being as public as
16 is consistent with Google's trade secrets. But what happens is
17 when you go beyond sort of a very shallow examination, you're in
18 the realm of trade secrets.

19 THE COURT: All right.

20 BY MR. SMURZYNSKI:

21 Q. Dr. Lehman, just to orient you, could you turn to the page
22 in this document that ends in .002?

23 A. Yes.

24 Q. Okay. And this refers to three pillars of ranking, and one
25 of them is listed as user interactions. And we've talked some

1 about that earlier today.

2 But what are those other two pillars of ranking?

3 A. So I guess by pillars here, these are sort of categories of
4 signal. So body is the content of a web page, all the texts,
5 the images, the title of the web page quite importantly, how
6 things are laid out, how things are colored, all that content.

7 Anchors are derived from links between web pages. So if
8 there is a hyperlink from page A to page B, then we would call
9 that link an anchor for page B, for the target page.

10 And when somebody has a hyperlink, they will often write --
11 well, there are a couple of things. First of all, if one person
12 links to another person's web page, that could be seen as a sort
13 of endorsement of the target page by the source.

14 But beyond that, hyperlinks often have texts, something
15 where you click the underlined words. And not always, but often
16 those underlined words are descriptive of the target page. Most
17 particularly, they often name the target page.

18 So for example, from anchors, it's easy to tell that
19 nytimes.com is The New York Times, because there are millions
20 and millions of links on the web that say NY Times and you click
21 on it and you get to nytimes.com.

22 Q. In order to evaluate anchors, is click user interaction
23 data used?

24 A. In order to evaluate anchors?

25 Q. Yes, in the system.

1 A. Let's see. So for -- maybe I'm not quite sure I'm
2 understanding the question. To generate the anchor signal,
3 that's just from links between web pages, and it doesn't involve
4 clicks. Is that the question?

5 Q. That was my question.

6 A. Okay.

7 Q. Okay. Now, you were also shown on page 004 a diagram that
8 referred to user interaction signals.

9 Do you see that?

10 A. Yes.

11 Q. And does this document capture all that's being done in
12 terms of how user interaction signals work with other signals
13 within Google to produce quality search results?

14 A. It looks like there's no mention of anchors or body or
15 other signals.

16 Q. And there's another document you were shown on page --
17 excuse me, Exhibit 204, ranking for research.

18 Do you recall looking at that document?

19 A. Yes.

20 Q. And you were shown at page 241, it had basic control flow.
21 So if you could go to that, ending in 241, Allen.

22 A. Yes.

23 Q. Does this capture the complexity of the Google search
24 engine?

25 A. No, not at all. I think the Google search engine was built

1 by first tens, then hundreds, then thousands of search engineers
2 over decades. It's just an immensely complex system.

3 MS. MURDOCK-PARK: Your Honor, I'm going to note that
4 again this was conducted in public. This is the second document
5 that he's using that we did in public session.

6 THE COURT: Understood.

7 Again, Counsel, I want to be mindful of time and all of
8 that, but if we're going to talk about things or ask him
9 questions that do not entrench on trade secrets specifically,
10 that should be done in a public session.

11 MR. SMURZYNSKI: I understand, Your Honor. I'm going
12 to go from the slide I just showed him to another slide in this
13 document that, without question -- I think there won't be any
14 dispute it involves trade secrets.

15 THE COURT: Okay.

16 MR. SMURZYNSKI: I'm going to have to ask a
17 preliminary question that I might draw some contention that it
18 could be done in open session. But it's impossible to do a
19 document and ask a question on one slide and then say hold that
20 thought, we're going to come back in 20 minutes, I'm going to
21 ask you about another slide on it.

22 Again, I understand the desire of the Court, but at some
23 point, it's almost impossible to do an examination of somebody,
24 like a ranking engineer, about something like Google ranking --

25 THE COURT: I understand. Let's move forward. The

1 last document, we didn't sort of descend into that particular --
2 as I thought you would be in that particular document. If
3 you're going to do that with others, then I understand the logic
4 of what you're saying. But again, just be mindful of how much
5 otherwise could be done in open session.

6 MR. SMURZYNSKI: I appreciate that, Your Honor.

7 BY MR. SMURZYNSKI:

8 Q. Dr. Lehman, do clicks present any problems as a signal in
9 ranking?

10 A. Yeah, they present many challenges.

11 Q. All right. And will you turn to the page that ends in 231
12 in this document.

13 THE COURT: I'm sorry. Can you repeat what exhibit
14 you're in?

15 MR. SMURZYNSKI: Certainly, Your Honor. We're in
16 UPX204, and it's the Bates number ending 231.

17 THE COURT: Okay.

18 BY MR. SMURZYNSKI:

19 Q. Dr. Lehman, at the top, the slide reads "some known live
20 traffic eval shortcomings."

21 Do you see that?

22 A. Yes.

23 Q. And how does that relate to the question of user data or
24 user interaction data?

25 A. So the chart is a little bit complex, but what it's

1 illustrating is one of the problems with using click data in
2 connection with ranking search results. It's a very strong
3 observation that people tend to click on lower-quality,
4 less-authoritative content than we would like to show on our
5 search engine. Our goal is to show -- when someone issues a
6 query, to give them information that's relevant and from
7 authoritative, reputable sources.

8 People tend not to click on those so much. So if we're
9 guided too much by clicks, our results would be of a lower
10 quality than we're targeting.

11 Q. And on this chart, on the left, there are various
12 indications of certain things and then the word "ablation."

13 What is that measuring?

14 A. So each line here refers to a ranking system or a
15 combination of ranking systems. And an ablation is an
16 experiment where you try turning off that system.

17 So what's happening here is roughly -- it's a little bit
18 complicated, but roughly, we're experimenting with turning off
19 these systems, and what we're observing is that that turning off
20 these systems brings up search results that get fewer clicks
21 than would otherwise be present.

22 Q. Dr. Lehman, you were read earlier a number of pages from
23 your deposition. I believe it was pages 397 to 398.

24 A. Yep.

25 Q. You were asked:

1 "Question: So is it your position that Navboost and QBST
2 are both capable of memorizing truly long-tail information on
3 relevance and user preference?

4 "Answer: Yes, Navboost and QBST are the two systems that
5 we would use for the -- rather, Navboost and QBST are the two
6 systems in our ranking system that -- most effective in
7 memorizing little facts about the world, little nuances of
8 language."

9 Did that observation make them the most important systems
10 or components in Google's ranking?

11 A. The most important, that observation, no.

12 Maybe to clarify a little bit, within Google Search, there
13 are systems like Navboost and QBST that are memorization
14 systems. So they have huge amounts of memory. And those
15 systems then become very good at memorizing little facts about
16 the world, because there are a lot of facts about the world.

17 There are also systems that involve considerable
18 computational complexity. That is, they do a lot of calculation
19 when they're processing data. Those systems tend to be
20 particularly good for understanding language, because language
21 has complex structure.

22 So if you want to understand facts about the world, you
23 need big memorization systems like QBST and Navboost. If you
24 want to understand the complexity of language, you need strong
25 computational systems.

1 And I realize that I kind of obscured that by letting in
2 this phrase "little nuances of language." Probably the other
3 part, the "little facts about the world," is what Navboost and
4 QBST are really good at.

5 Q. Okay. I'm going to direct your attention now to UPX19,
6 which was used in open session, though there are portions of
7 that document as well that have been marked highly confidential.
8 I'm going to ask you about some of those.

9 Do you have that in front of you, Dr. Lehman?

10 A. I'm sorry. Could you give me the number one more time?

11 Q. Certainly. It's UPX219.

12 A. Got it. Yes.

13 Q. This was a presentation you gave to the logging team in
14 COVID in 2020.

15 Do you recall testifying about that?

16 A. That's right, yes.

17 Q. And what was your motivation in giving that talk to the
18 logging team?

19 A. Well, it's the middle of COVID. The logging team does a
20 job that's not glamorous. They don't get to go on stage and
21 demonstrate hey, look at our cool logging technology, they don't
22 get any PR releases. But what they do is really important to
23 Google.

24 So they asked me to give a talk, and I tried to highlight
25 the really important role they play in the company.

1 Q. And did you use any examples with the logging team to help
2 explain how click data can be used at times by Google?

3 A. Yes. It looks like I gave an example on, I guess, page 401
4 and some parts afterward, yeah.

5 Q. If you could just generally describe that example to the
6 Court.

7 A. Yeah, so at home, we had a washing machine, and it's a
8 front loader. So if you completely close the door, it gets all
9 moldy. So there's this little magnet that keeps it part way
10 open.

11 And the magnet broke off of the washing machine. So I did
12 a search query like "washing machine door magnet broke off."
13 And the search results were really exceptionally good. It gave
14 me pages like from Amazon and elsewhere for the replacement
15 part.

16 And what I thought was kind of amazing is that I didn't
17 tell it anything else. I didn't tell it like what kind of
18 washing machine or anything of that nature.

19 MS. MURDOCK-PARK: Your Honor, I fail to see how with
20 respect to Mr. Lehman's washing machine this is confidential.

21 MR. SMURZYNSKI: We will get there with the next
22 question, Your Honor. This jumping up and objecting every three
23 questions on the basis of a desire to show this to the public,
24 it's just interrupting my exam, and at some point, it's got to
25 stop.

1 THE COURT: Let's move forward, please.

2 BY MR. SMURZYNSKI:

3 Q. Dr. Lehman, if you turn to page 407, do you have that in
4 front of you?

5 A. Yes.

6 Q. And Allen, if you could blow that up.

7 What's shown on page 407?

8 A. So this is showing the input and output of the QBST system,
9 which we discussed earlier. As a reminder, QBST takes as input
10 a search query, like "washing machine door magnet broke off,"
11 and returns a list of words and pairs of words that we believe
12 should be -- or that could be a system as determined should
13 appear prominently on web pages that are relevant to that search
14 query.

15 And I guess the striking thing to me and what I'm
16 highlighting here is that when I put in the query "washing
17 machine door magnet broke off," without specifying the washing
18 machine, without getting into any kind of details, it's showing
19 me the replacement part number.

20 Q. Was this the result of remote -- excuse me, rote
21 memorization by the Google Search system, or was there some
22 other work that went into creating tables such as this?

23 A. I guess the story of how those words appeared on the table,
24 I don't think I covered that in this presentation. It's -- I
25 will just sort of sketch the story. I won't define all the

1 terms, but just to kind of give a flavor of it.

2 So years ago, people started working on this thing called
3 the knowledge graph, and it tries to represent all the sort of
4 entities in the world. So that was a huge multi-year effort.

5 As that started to get going, some engineers in, I think,
6 Zurich had this idea that they could use it to improve the
7 ranking of search results. Again, without going into too much
8 detail, there's a technique called [REDACTED]
9 that they were able to use, and they created a special thing
10 that would sort of [REDACTED]

11 [REDACTED]
12 [REDACTED]. And so they built that up over a year. So
13 that launched.

14 Then some engineers in the Brazil office realized that that
15 system had some kind of systematic failures, and they had a new
16 approach for how to do it. And so they spent a year or so
17 building up this notion of salient terms and a new sort of
18 system based on salient terms.

19 And then some engineers in Mountain View a year or so later
20 realized that they could build another system on top of that,
21 and that is what became QBST.

22 So I guess no, it isn't like somebody just plugged in this
23 query and we just easily memorized this part number. In order
24 to get that part number into this table, it was a multi-year
25 engineering effort with multiple innovations by teams on three

1 continents.

2 Q. We can take that down.

3 Your Honor, may I approach?

4 Dr. Lehman, do you recall you mentioning various deep
5 learning models during the course of your exam?

6 A. Yes.

7 Q. And you have in front of you what's been marked as UPX255,
8 which is in evidence.

9 A. Okay.

10 Q. And the title of this document is "potential focus areas in
11 2020 and beyond."

12 Do you see that?

13 A. Yes.

14 Q. Are you familiar with this document?

15 A. Yes.

16 Q. What is this document?

17 A. I believe this was largely written by my close colleague
18 Sundeeep to discuss directions we might explore to make further
19 improvements to the search engine.

20 Q. Okay. And if you could turn forward to the page that ends
21 in 010, and at the bottom of the page, you will see there's a
22 section entitled "deep learned ranking and retrieval."

23 Do you see that?

24 A. Yes, I do.

25 Q. And as a reminder, this is -- this document is -- the focus

1 area is in 2020 and beyond.

2 What is being shown in the chart at the top of -- excuse
3 me, at the bottom of page 10, carrying over to page 11?

4 A. So I guess this is as of, I don't know, something like 2019
5 or 2020. This is showing four deep learning systems that had
6 been used in a ranking system.

7 Q. Okay. And when was rank brain introduced?

8 A. I think rank brain was introduced somewhere around 2014,
9 2015.

10 Q. What about deep rank next to it?

11 A. So deep rank came later. I think it's probably 2018 or
12 2019. Roughly, rank bank was built on sort of an older -- sort
13 of a first generation of deep neural network called a feed
14 forward or fully connected network.

15 Deep rank came later. It was built on a sort of second
16 generation language understanding -- second generation sort of
17 deep learning model incorporating something called BERT. BERT
18 was a breakthrough done by researchers at Google that radically
19 increased the ability of deep learning systems to understand
20 language.

21 Q. And in terms of the functionality they provide within the
22 Google search engine, how do rank brain and deep rank fit
23 together or not?

24 A. They're kind of -- they both sort of perform similar
25 functions in that once all the other ranking systems have

1 done -- have gone, or almost all, they look at the top [REDACTED]
2 search results, something like that, and adjust the scores based
3 on the sort of -- their neural networks.

4 So they're kind of a pair. The difference is one relies on
5 this older feed forward networks, which is rank brain, and deep
6 rank relies on sort of newer BERT technology.

7 Q. Okay. And if you go down the page, there's a row for
8 number of training examples.

9 Sorry, Allen. Keep up what you had.

10 How does rank brain, which was 2014-2015, compare to deep
11 rank in the 2018-2019 time period?

12 A. To clarify, [REDACTED] means order of or roughly [REDACTED]
13 training examples for rank brain. And for deep rank, the number
14 of training examples was about [REDACTED].
15 And I think this shows -- improvements in deep learning have
16 just vastly accelerated, but I think we're seeing here the early
17 ages of sort of the AI or deep learning revolution where you
18 have two systems doing similar things. They're providing
19 comparable improvements to search quality. And yet, as the sort
20 of technologies moved forward, only [REDACTED] of the training
21 examples are required.

22 Q. The next two columns are rank imbed and rank BERT. What's
23 the relationship between those systems?

24 A. So sort of as rank brain and deep rank kind of form a
25 natural pair, rank imbed and rank BERT are a natural pair. Rank

1 BERT was later renamed rank imbed version 2. So in terms of
2 their externally visual behavior, they're the same. And we
3 pulled out rank imbed and stuck in rank BERT, pulled out version
4 1 and put in version 2 subsequently, because it was
5 significantly higher performing.

6 Q. Even though it's using in that instance [REDACTED] of the
7 amount of data, what's the performance?

8 A. So yeah, I think for me this is really striking. When we
9 went from rank imbed, we were using about [REDACTED] training
10 samples. And when we switched to rank BERT, we were able to get
11 significantly larger contributions to search quality with
12 [REDACTED] the training data.

13 Because the BERT technology is so well adapted to working
14 with language -- actually, the inventor of BERT made rank BERT.
15 It was transformational. Once it appeared, so many variants
16 appeared all across the tech industry because it was just so far
17 ahead of everything else.

18 So these are examples of systems that are really good at
19 understanding the complexity of language, grammatical structure,
20 things like that. So they do a lot of computation. They may
21 not be so good at memorizing facts, but they're really good at
22 understanding language.

23 MR. SMURZYNSKI: Your Honor, I have maybe ten minutes
24 more of questions in closed session. I realize it's 5:00.
25 However Your Honor wants to approach it, of course, we will do

1 whatever you want, but I just offer that observation.

2 THE COURT: Well, how much do you think in redirect in
3 closed session and then how much do you think you will have in
4 an open session?

5 MR. SMURZYNSKI: I think maybe 10, 15 minutes in
6 closed and maybe 10 in open. And I'm sort of -- I'm trying to
7 adjust on the fly here to make sure that we push as much into
8 open as we can, but there's -- sometimes there's a flow.

9 THE COURT: Sure.

10 Counsel, what do you think in terms of your redirect on
11 both fronts?

12 MS. MURDOCK-PARK: Limited redirect, but I will need a
13 moment to consult.

14 THE COURT: Look, I do need to -- I know this will,
15 unfortunately, require Dr. Lehman to remain overnight, but I
16 really do have to end today for our courtroom staff, and then
17 we've got a couple of things we need to just close up.

18 So we'll start tomorrow at 9:30. We will begin in a closed
19 session. And then we will come out of the closed session once
20 the redirect in the closed is concluded.

21 All right. Dr. Lehman, apologies. We're going to have to
22 keep you here overnight. We will ask you to be back here so
23 that we can begin at 9:30, and I ask you not to discuss your
24 testimony overnight with anyone.

25 THE WITNESS: Okay. Thank you.

1 THE COURT: Why don't we go ahead and reopen the
2 courtroom, and I will ask Dr. Lehman to step out.

3 (End of sealed proceedings.)

4 THE COURT: In terms of open items, we've got the
5 Exhibit 38 issue.

6 Does Google wish to be heard on the admissibility of 38,
7 Counsel?

8 MR. BENNETT: Yes, Your Honor.

9 So the government has moved for the admission of Exhibit
10 38, and they bear the burden, obviously, by a preponderance of
11 the evidence to show that it has a basis to come in. They've
12 moved only on the grounds of Rule 801(d)(2), and they've
13 emphasized (d)(2)(D) as the basis for that.

14 Their argument, Your Honor, is, I guess, intended to show
15 that Exhibit 38, which Your Honor has seen and heard much of, is
16 a statement by Google. Google is the party opponent; Mr. Roszak
17 is not the party opponent.

18 And to make that argument, they really focused, the
19 government focuses on half of 801(d)(2)(D). They focus on the
20 temporal aspect, was -- and this is their brief that was filed,
21 was the witness employed by Google at the time he made the
22 statement. The statement was from 2017. As Your Honor heard,
23 Mr. Roszak was, in fact, employed by the government at that
24 point -- not by the government, sorry, by Google at that point.
25 He was a director. As Your Honor heard, there were dozens and

1 dozens of groups in finance. He was in one of them. And so
2 that's the standard that the government puts forth.

3 THE COURT: Let's sort of cut to the heart of it. Why
4 do you think his statements don't fall within the scope of the
5 agency or the employee relationship?

6 MR. BENNETT: So I would submit that some of the
7 statements come close. The one that the government focused on,
8 counting the bullet points down, Your Honor, it would be the
9 fifth and sixth bullet points. Those ones come close to what he
10 focused on at Google. The other statements are clearly not
11 within his wheelhouse. The statements about -- they're really
12 macroeconomic statements, microeconomic statements, statements
13 about parts of the business where he was not involved.

14 So it's like a lot of other documents Your Honor has
15 seen --

16 THE COURT: Doesn't that ultimately go to the weight
17 of the questioning -- the weight of what he's saying here?

18 I mean, look, he was the director of finance at the time he
19 did this. It was a Google-sponsored training. And he made the
20 decision to provide a -- to give a training speech that involved
21 the area of his knowledge and understanding.

22 Again, I understand it's probably at the outer edges here.
23 It's not at the core as we would see with a communication
24 between colleagues in the context of financial -- you know,
25 discussion about finance. But the standard is pretty lax,

1 pretty broad. And the cases that you all have cited don't quite
2 go as far as I think you've suggested they do.

3 MR. BENNETT: Your Honor, I think the standard is
4 probably most clearly set out in the *Aliotta* case that the
5 government cites. That's a case where someone who was employed
6 as an accident reconstructionist at a train company made some
7 comments in a deposition about a train accident.

8 And the Court there distinguished those factual statements
9 about a train accident by a train accident reconstructionist
10 from -- and they juxtaposed that with what if he had offered
11 conclusions about this particular train accident. And they said
12 he was hired to do the former; he wasn't hired to do the latter.

13 There's no basis to think that Mr. Roszak was hired by
14 Google, employed by Google to make the comments in the first
15 four and the sixth bullet point. That's why the government got
16 up -- I'm sorry, the first four and the seventh bullet point.
17 That's why the government got up and asked about the fifth and
18 sixth, because they know they had to find some sort of hook.

19 So this to me looks like a document that the Court has seen
20 many times where there is hearsay and nonhearsay. And the best
21 argument the government can make is that there's some
22 nonhearsay.

23 Now, we would ask Your Honor, because those parts that the
24 government talked about were talked about in closed session,
25 those are confidential. And if the document comes in, it should

1 be redacted, as we have done in many documents, redacting out
2 the things that are -- I would say irrelevant, but I know Your
3 Honor wants to consider the weight, are clearly hearsay.

4 To go back to something Your Honor said a second ago, I
5 think Your Honor said that this was presented by him, he was
6 ordered to give a training. He wasn't giving a training. He
7 was taking a training.

8 THE COURT: No, I understand.

9 MR. BENNETT: And I think what distinguishes this from
10 the government's lead case is that in that case the declarant,
11 the witness was being deposed. He was testifying about things
12 that were, at least he was asserting, true, related to his job
13 and they were true.

14 Here, you've got something that the witness has said was
15 not true, wasn't intended to be true. He was cosplaying Gordon
16 Gekko. He was saying things that he didn't think were true
17 because it was a part of the training to do that.

18 If they had done a skit, the government certainly wouldn't
19 come in and offer that for the truth of whatever role he was
20 playing. This is the equivalent of that.

21 So the *Aliotta* case, I think, really is guiding and that
22 there are -- there's a dividing line between things that are in
23 an employee's wheelhouse, what they are hired, what they are
24 employed to do, and things that are not.

25 And if Your Honor looks at --

1 THE COURT: What the Seventh Circuit said in *Aliotta*,
2 315 F.3d 756, is that the only requirement -- and this was, I
3 think, in a Title VII case, but the only requirement is that the
4 subject matter of the admission match the subject matter of the
5 employee's job description.

6 MR. BENNETT: And then the Court goes on to say that
7 even though concluding about that accident was the subject
8 matter of his employment, they said that would be excluded
9 because he wasn't employed to make conclusions about that
10 accident. He was employed to know about accidents generally.

11 Here, we're one step removed from that. Mr. Roszak is not
12 an economist. Your Honor, I think, has heard from economists
13 already in this case. He's not an economist. He's not a
14 strategy person or a policy person. He's someone who does
15 financial modeling in a very narrow area of the business. In
16 fact, he draws that out elsewhere in this memo, that he's
17 talking about other parts of the business, not the part where he
18 works.

19 So this just --

20 THE COURT: Here's the difficulty of where I find
21 myself, which is that to the extent this is coming in, it does
22 not contain anything that would qualify as confidential. It
23 doesn't contain a trade secret. It doesn't contain any
24 financial information that has been sort of in the wheelhouse of
25 what is confidential.

1 Now I'm in the position where the testimony that
2 accompanied this document was done in a closed session. So all
3 context is lost if this document is admitted and then finds its
4 way into the public space, because in theory it is now a public
5 document if I admit it.

6 So I'm in a little bit of a pickle, in a jam here because
7 of how we have teed this up.

8 So let me just go ahead and rule. We've got to bring
9 things to a close here today.

10 MR. BENNETT: If I could just --

11 THE COURT: Thank you.

12 So look, the relevant provision here is 801(d) (2) (D), and I
13 want to just read a few excerpts from the *Wright & Miller*
14 *Treatise*, which is paragraphs 67 to 76, that covers it. This is
15 30B, Federal Practice and Procedure, in evidence.

16 Counsel, you can have a seat.

17 MR. BENNETT: Thank you, Your Honor.

18 THE COURT: *Wright & Miller* states, "The exemption
19 requires neither adoption nor ratification of the statements by
20 the party. All that is required is that the agent or employee
21 made the statements during and within the scope of the agency or
22 employee relationship.

23 "With respect to the subject matter, Rule 801(d) (2) (D)
24 abandoned the common law requirement that the agent must have
25 authority to make statements in a given area. It is no longer

1 the law that a qualifying statement must be shown to have been
2 made by the employee at the insistence of her employer.

3 "Instead, 801(d) (2) takes a broader view, allowing
4 statements to be admitted over a hearsay objection whenever an
5 employee or agent speaks on any matter within the scope of his
6 agency or employment during the existence of that relationship.

7 "By virtue of this requirement, qualifying statements are
8 to some degree limited to those the employee is most likely to
9 be knowledgeable about. Qualifying statements will arise when
10 an employee or agent speaks about matters that are or were
11 within her job responsibilities."

12 And then finally, the treatise says, "Outside the context
13 of employee complaints, Courts typically take an unabashedly
14 broad view of Rule 801(d) (2) (D) 's application, emphasizing that
15 qualifying statements must simply touch on matters that fall
16 within the employee or agent's general duties and
17 responsibilities.

18 "Stated another way, an employee need only be performing
19 the duties of his employment when he comes in contact with the
20 particular facts at issue. Importantly, the statements must
21 still concern matters related to the speaker's employment."

22 And so I think by those standards this comes in as an
23 801(d) (2) (B) -- excuse me, 801(d) (2) (D) statement. I will grant
24 it's on the outer perimeter here, but it nevertheless meets all
25 of the requirements. Here you have somebody who is the director

1 of finance, who albeit in a training session made the decision
2 to make a mock presentation, understanding that's what it is,
3 mock presentation about things he understands about the
4 business, about the fundamentals of the business.

5 And they're really broad-stroke statements, obviously, but
6 he doesn't need to be an economist to make statements like this.
7 He has advanced degrees in finance, I assume, or an MBA and is
8 the head of Google Finance. These are not statements that would
9 fall outside his wheelhouse, so to speak, or outside his areas.

10 I mean, all the other indicia establish that this was a
11 statement made during the course of his employment. It was a
12 Google-sponsored training. It was done off-site, presumably
13 sponsored by Google. It was done on his computer at work or his
14 work-issued computer. And so all of those indicia establish
15 what is required to meet the elements of the rule.

16 I will just note the handful of cases that Google has cited
17 starting with *Rowell* from the Eleventh Circuit, 2005, is
18 inapposite because in that case the statement was deemed not
19 admissible because there was no evidence that the employee was
20 involved in the decision of how to structure the reduction in
21 force or how to design the performance evaluations used during
22 the nonvoluntary stage of the reduction. There's also no
23 evidence that the employee had received any information from the
24 employer's upper management to indicate that age was to play a
25 factor in these decisions.

1 As I said, this is different. This is a topic about which
2 he can speak authoritatively based upon his experience. I can't
3 remember. I think he said he had been at Google for
4 approximately ten years by the time this training happened, the
5 entire time in finance.

6 The Seventh Circuit's decision in *Young* notes that the
7 exception or the carve-out from the rule embodies broader
8 exception than the prerules. The rule simply requires that the
9 statement be made by an individual who is an agent and that the
10 statement be made during a period of the agency and that the
11 matter be within the subject matter of the agency.

12 And the reason in that case the statement was excluded is
13 because the statement was made in a resignation letter, and the
14 Court concluded that essentially the employee was acting as an
15 adversary and no longer inhibited by his relationship with the
16 principal from making erroneous or underhanded comments that
17 could harm the principal.

18 And then finally, the *Blackburn* decision from the Third
19 Circuit, 1999, in which the Court excluded the statement or
20 affirmed the exclusion of the statement where there was no
21 indication that the employee was speaking for UPS on a matter
22 within the scope of his agency employment such that the employee
23 could testify about what he had been told.

24 And then finally is the *Mister* case from the Seventh
25 Circuit, the case involved a report. What the Court said there

1 is not that the report was inadmissible under 801(d)(2)(D), even
2 if it contained an absurd statement, but rather, it was
3 excludible under Rule 403. So even though it met the
4 requirements of 802(d)(2)(D), the circuit said it was properly
5 excluded under Rule 403.

6 So again, I think this meets the requirements of
7 801(d)(2)(D). In terms of the 403 balancing, look, I understand
8 it's somewhat embarrassing to the witness, but it's not
9 substantially -- the probative value -- the likelihood of
10 confusing me or creating embarrassment for the witness and for
11 Google is not substantially outweighed by the probative value of
12 what's in here.

13 He's testified that he doesn't believe a lot of what he's
14 said. He's provided context for the way in which he said it.
15 So all of that will go to the weight of the ultimate
16 consideration of the statement.

17 But in terms of its admissibility, I think it meets the
18 four corners of the rule. And so I will admit it.

19 So UPX38 can be admitted.

20 (UPX38 received into evidence.)

21 THE COURT: I will leave it to you all to discuss
22 whether it ought to be redacted in any way. Talk about that
23 amongst yourselves, and we can deal with that overnight. Okay?
24 Or in the morning.

25 One last housekeeping matter. I think counsel had

1 requested at the -- or in fact, I know you all had requested
2 that we create essentially a Webinar link so that there would be
3 a Webinar available to counsel of record, both for the parties
4 and third parties to access the proceedings. We allowed that.
5 At the time it was something that was permitted by judicial
6 policy. That policy expires as of tomorrow.

7 And it's not clear to me whether continuing that Webinar
8 line is something I can do under the current policy. I can
9 provide everyone with what the policy changes are, and we can
10 talk about it in the morning, but I just wanted everybody to be
11 aware it's something that's been brought to my attention and
12 something I need to consider before the end of the day tomorrow.

13 So let me give these two documents to Mr. Douyon, who can
14 make photocopies of them for you, and you can think about it
15 overnight, and we can talk about it in the morning.

16 Al right. Is there anything else we need to talk about?

17 MR. CAVANAUGH: No, Your Honor.

18 MR. DINTZER: Not from the DOJ plaintiffs, Your Honor.

19 MR. SCHMIDTLEIN: No, thank you, Your Honor.

20 THE COURT: Thank you all very much. Please do not
21 wait for me. And we will see you tomorrow.

22 (Proceedings adjourned at 5:18 p.m.)
23
24
25

1 CERTIFICATE OF OFFICIAL COURT REPORTER

2
3 I, Sara A. Wick, certify that the foregoing is a
4 correct transcript from the record of proceedings in the
5 above-entitled matter.
6

7
8 /s/ Sara A. Wick

September 20, 2023

9 SIGNATURE OF COURT REPORTER

DATE

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25