

EXHIBIT C

IN THE UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

ELON MUSK, et al.,

Plaintiffs,

v.

SAMUEL ALTMAN, et al.,

Defendants.

Case No. 4:24-cv-04722-YGR

VIDEO DEPOSITION OF

Stuart Russell, OBE FRS

December 2, 2025

Washington, D.C.

LEAD: Adam Tanne, Esquire

FIRM: Wachtell Lipton Rosen & Katz

FINAL COPY - CONFIDENTIAL

JANE ROSE REPORTING - 1-800-825-3341

A P P E A R A N C E S

ATTORNEYS FOR PLAINTIFFS

Eric Nitz, Esquire

Robert Kry, Esquire

MOLOLAMKEN

600 New Hampshire Avenue, Northwest

Washington, D.C. 20037

(202) 556-2000

ATTORNEYS FOR DEFENDANTS

Nathaniel Cullerton, Esquire

Adam Tanne, Esquire

Ioannis Drivas, Esquire

WACHTEL LIPTON ROSEN & KATZ

51 W 52nd Street

New York, New York 10019

(212) 403-1000

ATTORNEY FOR MICROSOFT

Nisha Patel Gupta, Esquire

DECHERT, LLP

633 W Fifth Street, Suite 4900

Los Angeles, California 90071

(213) 808-5700

JANE ROSE REPORTING

74 Fifth Avenue

New York, New York 10011

1-800-825-3341

Christina Hotsko, Court Reporter

Jonathan Rasson, Videographer

TABLE OF CONTENTS

WITNESS: STUART RUSSELL

EXAMINATION BY:	PAGE
Counsel for OpenAI Defendants.....	05
Counsel for Plaintiffs.....	289
Reporter Certificate.....	293
Index of Exhibits.....	294
Notice to Read and Sign.....	296

1 A. Nope.

2 MR. CULLERTON: This would be a good
3 place to take five minutes, if that's all right.

4 MR. NITZ: Great.

5 VIDEO TECHNICIAN: The time is 10:12, and
6 we are off the record.

7 (A recess was taken.)

8 VIDEO TECHNICIAN: The time is 10:23, and
9 we are back on the record.

10 BY MR. CULLERTON:

11 Q. Professor Russell, you are not offering
12 an opinion concerning any alleged safety
13 commitments made by OpenAI to Mr. Musk; is that
14 accurate?

15 A. That's correct.

16 Q. And you're not offering an opinion on
17 whether OpenAI's current or past safety practices
18 are consistent with any alleged commitments
19 OpenAI made to Mr. Musk, correct?

20 A. Correct, yes. I'm not aware of what
21 commitments were made.

22 MR. NITZ: I'm sorry to interrupt. It
23 looks like our realtime feed is not updating. If
24 we could just take a minute to go off the record.

25 VIDEO TECHNICIAN: The time is 10:23, and

1 we are off the record.

2 (Discussion off the record.)

3 VIDEO TECHNICIAN: The time is 10:25, and

4 we are back on the record.

5 BY MR. CULLERTON:

6 Q. And Professor Russell, you're not
7 offering an opinion on whether OpenAI's safety
8 practices are better or worse than the safety
9 practices of other AI companies, now or in the
10 past; is that correct?

11 A. That's correct.

12 Q. Okay. And you -- in forming your
13 opinions and preparing your report, you didn't
14 compare the relative strengths and weaknesses of
15 safety practices employed by different AI
16 companies; is that correct?

17 A. In preparing my report, that's correct.

18 Q. And you're not offering an opinion on the
19 relative strengths or weaknesses of safety
20 practices at different AI companies, correct?

21 A. Correct. Yeah. If asked, I can give my
22 opinion, but I have not put that in my report.

23 Q. All right. And you're not offering an
24 opinion on risks caused by any alleged past
25 safety failures by OpenAI, correct?

1 A. So I'm trying to think. I have examples
2 in the report of the kinds of -- some of the
3 kinds of safety failures that are occurring. And
4 I don't recall -- I think the -- there's a
5 transcript, but I think that's from Claude, which
6 is an Anthropic product.

7 Q. Okay.

8 A. I don't recall specifically if I mention
9 a safety failure by OpenAI.

10 Q. Okay.

11 A. I think, by implication, the general
12 views that I give on safety failures would
13 include OpenAI's products, but I'm not making a
14 specific claim about OpenAI's failures.

15 Q. To the extent you're making any claim
16 about safety, that claim would apply equally to
17 AI companies generally; is that fair?

18 A. Yeah, it could, depending on exactly
19 which claim I was making, yes.

20 Q. And which of the four -- and you can look
21 at the report -- opinions summarized in your
22 report deal with alleged safety failings by AI
23 companies, if any?

24 A. So I think two I would say -- by
25 implication, my view is that there has not, in

1 general, been sufficient investment in safety.
2 So although it's stated as a conditional, i.e.,
3 if there isn't further investment in progress --
4 so I guess, by implication, I'm saying there
5 hasn't been sufficient investment in safety
6 overall.

7 And then in opinion 3 -- right. The
8 claim is that there are -- there are strong
9 incentives to develop capabilities quickly and
10 that those incentives are in conflict with what I
11 view as a requirement to ensure safety.

12 Q. And we'll get into those in more detail.

13 But again, at least as you just stated
14 it, that potential conflict would apply equally
15 to any of the leading AI companies?

16 A. Yes, it would.

17 Q. Okay. Are you going to offer any --
18 offering any opinion in this case on whether any
19 existing OpenAI product is safe or unsafe?

20 A. I am not.

21 Q. Okay. And are you offering an opinion in
22 this case on whether any historical OpenAI
23 product is safe or unsafe?

24 A. I am not, I should say, beyond the
25 general opinion that I think all products based

1 on large language models have an intrinsic lack
2 of safety due to the way they're constructed.

3 Q. And you are not offering an opinion in
4 this case on OpenAI's relationship with
5 Microsoft; is that accurate?

6 A. That's accurate.

7 Q. Okay. And you're not offering an opinion
8 as to whether that relationship was good or bad
9 from a safety perspective, correct?

10 A. Correct.

11 Q. And you're not offering an opinion
12 concerning OpenAI's corporate governance,
13 correct?

14 A. Correct.

15 Q. Okay. You're not offering an opinion on
16 OpenAI's nonprofit entity, correct?

17 MR. NITZ: Objection.

18 THE WITNESS: Okay.

19 BY MR. CULLERTON:

20 Q. You can answer.

21 A. Correct. There's nothing in my report
22 that discusses the corporate structure of OpenAI.

23 Q. Okay. And so, therefore, I take it,
24 you're not offering an opinion on whether
25 OpenAI's past or current corporate structure

1 influenced its safety practices in any way,
2 correct?

3 A. So the general view that competition
4 among companies is leading to -- or tends to
5 reduce the focus on safety would apply. And I
6 suppose, by implication, if OpenAI transitions
7 from the control of a nonprofit board to some
8 other structure, that would -- that would, I
9 think, tend to eliminate one restraint on
10 OpenAI's conduct and its focus on safety.

11 Q. Okay.

12 A. That's my understanding of how --

13 Q. But that's not an opinion you're --

14 A. -- the corporate structure --

15 Q. That's not an opinion you're offering in
16 your --

17 A. In my report I'm not specifically
18 addressing that. I'm just saying it's an
19 implication of the general view.

20 Q. And you're not offering an opinion on
21 whether OpenAI's safety practices are consistent
22 with its nonprofit mission, correct?

23 A. Correct.

24 Q. Okay. And you're not offering an opinion
25 on OpenAI's recently announced restructuring,

1 correct?

2 A. Correct.

3 Q. And you're not offering an opinion as to
4 whether that restructuring would be good or bad
5 from a safety perspective, correct?

6 A. Correct. Other than, as I explained,
7 from the general view that competition is tending
8 to reduce the focus on safety, and so that
9 transition might therefore be deleterious to
10 their focus on safety.

11 Q. Did you study the structure of the
12 restructuring at all in connection with forming
13 your opinions in this case?

14 A. Not in any great detail, and it wasn't
15 relevant to the opinions I was asked to give.

16 Q. Okay. So you were not asked to provide
17 an opinion as to whether OpenAI's restructuring
18 would be good or bad from a safety perspective;
19 is that right?

20 A. Correct.

21 Q. Okay. And are you offering an opinion on
22 whether artificial general intelligence should be
23 developed or not by anyone?

24 A. Not as such, no. I'm of the view, and I
25 think it's part of the opinion, that it should

1 only be developed if there are adequate safety
2 measures accompanying it.

3 Q. Okay. And you're not offering an opinion
4 that OpenAI specifically, as opposed to any other
5 AI company, has prioritized profit over safety,
6 correct?

7 A. Correct.

8 Q. Is it your opinion that LLM-based AI
9 systems are inherently unsafe and cannot be made
10 safe?

11 A. It is. Yes.

12 Q. Is it your opinion that current LLM-based
13 AI systems pose a threat to humanity?

14 A. No.

15 Q. Okay. So the opinion is that they could,
16 if they were -- if the development path continued
17 along what you discuss here towards AGI --

18 A. Yes.

19 Q. -- is that fair? Okay.

20 When did you first meet Elon Musk,
21 Professor Russell?

22 A. It was January of 2015.

23 Q. Okay. And how would you describe your
24 relationship with him at that time?

25 A. So this was at a meeting organized by

1 the remarkable speed with which the companies are
2 able to develop and release significantly
3 improved versions of their software.

4 Q. Let's go back to paragraph 45.

5 In the last sentence you write, These
6 considerations create race conditions that may
7 lead to safety being a secondary consideration.

8 Do you see that?

9 A. Yes.

10 Q. Okay. And the "these considerations"
11 being referred to there, just so I understand, my
12 understanding is, one, what you say are the
13 enormous economic value of AGI --

14 A. Yeah.

15 Q. -- and two, first-mover advantage.

16 A. Yeah.

17 Q. Okay. And you write -- and for the
18 proposition that there are race conditions that
19 may lead to safety being a secondary
20 consideration, you cite, in footnote 27,
21 The Economist.

22 Do you see that?

23 A. Yes.

24 Q. Okay. And did you review that Economist
25 article?

1 A. Briefly. It was really -- I'm not saying
2 The Economist article proves my case, but more
3 pointing to that article as showing that this is
4 a -- you know, a commonly held opinion among
5 people looking at what's going on.

6 Q. Do you know The Economist article is
7 anonymous?

8 A. I think that's normal for Economist
9 articles --

10 Q. Okay. But you don't know who the author
11 is?

12 A. Correct. I don't know who the author is.

13 Q. Were you comfortable relying on an
14 anonymous article in a news publication to
15 support your expert opinion in this case?

16 A. As I said, I'm not relying on that
17 article to support the argument, more to show
18 that that argument is something that is already
19 well known and widely talked about.

20 Q. Okay. Well, your opinion is that these
21 considerations, the two we just discussed, create
22 race conditions, right?

23 A. Yeah.

24 Q. Okay. And --

25 A. Meaning that the --

1 Q. Other than The Economist article, which I
2 understand you now to be saying you're not
3 relying on in support of that opinion, what are
4 you relying on in support of the opinion that
5 there are race conditions?

6 A. So race conditions here meaning that the
7 companies feel a very strong incentive to move
8 the capabilities forward as fast as they can, to
9 release systems with better capabilities than
10 their competitors as soon as possible, that they
11 feel threatened by potential advances that other
12 companies might make.

13 So I think those conditions are well
14 known. I don't see that they need detailed
15 arguments in support that that's the situation.

16 Q. Okay. And you say the companies feel
17 this.

18 In forming your opinions and drafting
19 your expert reports, did you poll AI companies to
20 assess their views of the competitive dynamics in
21 the industry?

22 A. Did I poll the companies?

23 Q. Yes.

24 A. No. I have a lot of conversations with a
25 lot of people, and I've not heard anyone say

1 anything to contradict this view, including, for
2 example, the CEO of Google DeepMind, who
3 expressed very -- very similar concerns, that
4 they were in a race that they couldn't pull out
5 of.

6 Q. And when did that conversation occur?

7 A. That would have been late June of this
8 year.

9 Q. But you're not relying on that
10 conversation in forming your opinion in this
11 case, right?

12 A. Along with many, many other
13 conversations, it has contributed to my
14 understanding of the situation as it exists
15 today.

16 Q. Okay. Have you talked to every CEO of an
17 AI company in forming your opinion that there are
18 race dynamics?

19 A. No, I have not talked to the CEO --

20 Q. Okay.

21 A. -- of every AI company.

22 Q. Okay. Have you done any analysis of
23 market share of the various AI companies in
24 forming your opinions in this case?

25 A. I have not studied the details of market

1 share over time.

2 Q. Okay.

3 A. And --

4 Q. None of the --

5 A. -- I would also say that the notion of
6 market share is not particularly well defined.

7 For example, there are a lot of AI services that
8 are available for free. And do you want to count
9 those as part of market share or not? And so on.

10 But I think the idea that these -- so I
11 think the three leading companies, and then there
12 are others -- Perplexity and xAI and then a few
13 smaller players like Cohere -- but certainly the
14 three leading companies view the competition as
15 intense.

16 And you can -- you know, there's a lot of
17 anecdotal evidence. You can see, for example,
18 Meta offering hundred-million-dollar sign-up
19 packages to steal researchers from other
20 companies. You know, that's the kind of thing
21 that happens when there is this intense
22 competitive dynamics.

23 Q. Okay. But your opinion is that there is
24 a race to AGI that may lead the companies to
25 deprioritize safety, right?

1 A. That is my opinion. Yes.

2 Q. Okay. And I was just trying to
3 understand the bases for that opinion since it
4 doesn't appear to be disclosed in your report.

5 MR. NITZ: Objection.

6 BY MR. CULLERTON:

7 Q. You cite The Economist article, correct?
8 Correct?

9 A. I do cite The Economist article.
10 Correct.

11 Q. Okay. Is there any other source
12 disclosed in your expert report that you contend
13 cite -- supports the opinion that you are
14 offering, that there are race dynamics in the AI
15 industry that may lead companies to deprioritize
16 safety?

17 A. I would have to check, but I don't think
18 I cite any other sources for the company
19 dynamics. There are sources cited for the
20 geopolitical race.

21 Q. Focusing just on the company dynamics.

22 A. Just on the company dynamics?

23 Q. Yes.

24 A. I don't think I cite any other sources
25 for that.

1 Q. Okay. Are you aware that there's been
2 published academic work on race dynamics in the
3 AI space? Including by Han, Cimpeanu, others?

4 A. I don't think I've read those papers.

5 Q. Okay. Did you consider any academic work
6 on race dynamics in the AI industry in forming
7 your opinions in this case?

8 A. I did not consider academic papers on
9 that topic.

10 Q. Let's take a look at The Economist
11 article.

12 MR. CULLERTON: This is going to be
13 Exhibit 7.

14 (Russell Deposition Exhibit 7 marked for
15 identification and attached to the
16 transcript.)

17 BY MR. CULLERTON:

18 Q. Do you recognize this, Professor Russell,
19 as the article cited in your report?

20 A. Yes, I do.

21 Q. All right. And I'll just direct you to
22 the very last page of the article, the
23 paragraph that is continuing from the prior page,
24 the sentence beginning, "Anyway."

25 Do you see that?

1 A. Uh-huh.

2 Q. Anyway, for every expert predicting an AI
3 apocalypse, there's another who insists there is
4 nothing to worry about.

5 Do you see that?

6 A. I do. Yes.

7 Q. Okay. Do you agree with that
8 characterization by The Economist?

9 A. No, not really.

10 Q. Okay.

11 A. I think there -- it's a figure of speech.
12 I don't think they mean literally one-for-one.

13 Q. Do you agree that there are serious AI
14 experts who dispute the severity of the potential
15 risks from AI?

16 A. Yes. I agree there are some.

17 Q. And you didn't reference any of those in
18 your expert report, correct?

19 A. Correct. My job is to give arguments for
20 the propositions rather than against my own
21 propositions.

22 I should add that I have a whole chapter
23 in my book which considers the arguments made for
24 not paying attention to these risks and addresses
25 each of those arguments. And in the six years

1 since the book has been published, not one of
2 those refutations has been disputed successfully
3 as far as I know.

4 Q. Thank you.

5 MR. CULLERTON: Let's take five minutes
6 here. Thanks.

7 VIDEO TECHNICIAN: The time is 1507, and
8 we are off the record.

9 (A recess was taken.)

10 VIDEO TECHNICIAN: The time is 1522, and
11 we are back on the record.

12 BY MR. CULLERTON:

13 Q. Professor Russell, just looking back at
14 paragraph 45, in which you state your opinion
15 regarding race conditions, the last sentence
16 states, These considerations create race
17 conditions that may lead to safety being a
18 secondary consideration.

19 Fair to say, then, that you are not
20 offering an opinion that safety is currently a
21 secondary consideration, but that it may become
22 so as a result of the race conditions you
23 identify?

24 A. I am not offering an opinion either way
25 beyond the points made that the safety measures

1 that have been developed so far are found wanting
2 in many cases.

3 Q. Okay. But your opinion is not that any
4 of those failings have been a result of the race
5 conditions that you're identifying in
6 paragraph 45.

7 A. My expert opinion, this document does not
8 state that.

9 Q. Okay. Now do you agree that AI companies
10 also have significant economic incentives to
11 ensure that their products are safe?

12 A. They have some. I think I think the fact
13 that, generally, they disclaim liability for the
14 effects of their AI systems means that their --
15 that significantly reduces the incentive to
16 ensure that their products are safe.

17 Q. You're not offering an opinion in this
18 case about any alleged disclaimers of liability,
19 correct?

20 A. I am aware that at least some of the
21 companies disclaim liability. For example, I
22 have looked at the licenses of Cohere and
23 Microsoft, both of which largely disclaim
24 liability. And I believe this is common practice
25 not just among AI companies but among software

1 companies in general.

2 Q. Okay. Have you cited any of these
3 licensing agreements in your expert report?

4 A. No, I have not.

5 Q. Okay. Did you consider these license
6 agreements in connection with forming your
7 opinions in this case?

8 A. With the opinions of the report, no --

9 Q. Okay.

10 A. -- because the opinions of the report are
11 not contingent on that aspect.

12 Q. Now, you gave the Reith Lectures on BBC
13 in 2021, right?

14 A. I did, yes.

15 Q. Okay. Do you recall during those
16 lectures saying that, I believe companies will
17 have a very strong economic incentive to adopt
18 this new approach as soon as it's feasible,
19 referring to the human-compatible approach that
20 you were proposing?

21 A. Yes.

22 Q. Okay. And given a number of talks where
23 you use an analogy about a robot cooking a cat at
24 home.

25 Do you know what I'm referring to?

1 Do you see that?

2 A. I do. Yes.

3 Q. Okay. You're not offering your own
4 opinion on the percentage probability of
5 extinction risk, correct?

6 A. No, I am not in this case.

7 Q. How come?

8 MR. NITZ: I'll object and instruct you
9 not to get into, you know, attorney work product
10 and, you know, strategy discussions. But you can
11 otherwise answer outside that.

12 THE WITNESS: Okay. So I think my risk
13 calculation or risk estimate works differently
14 from the CEOs, and I -- so I've presented it
15 publicly so I can discuss it.

16 I think the two main branches are, do we
17 succeed in developing something like AGI that
18 could present this type of risk in the relatively
19 near future or not? And if we don't, then I
20 think there's a substantial likelihood that the,
21 quote, AI bubble would burst. So if we're not
22 delivering that type of capability, then the
23 investments that have been made and are going to
24 be made probably won't return the necessary
25 value. And the bubble bursting, meaning as

1 happened in the dot-com bubble bursting and the
2 previous AI winter that happened in the
3 mid-1980s.

4 So that would introduce a potential delay
5 of at least a decade because investment levels
6 would collapse. In my experience, interest among
7 students would collapse, whereas right now
8 interest in AI is intense; almost all applicants
9 to our department want to study AI. So that
10 would collapse.

11 And so it would give us another decade
12 during which we might be able to make the
13 necessary advances in safety.

14 And probably as happened after the bubble
15 burst in the mid-'80s, the field would focus much
16 more on the real science and engineering, and
17 possibly divert the direction of technology into
18 a direction where we can actually make the
19 necessary quantitative guarantees of safety.

20 So that's one branch. Then the other
21 branch is that some further breakthroughs happen
22 that do lead to something like AGI in the
23 relatively near term. And then I think the risk
24 would be quite high that we would experience some
25 significant negative outcomes.

1 And you can roughly divide those into
2 irreversible catastrophe or a reversible
3 catastrophe that would be big enough to change
4 both corporate -- the direction of technology
5 that corporations are developing and the views of
6 government about the need to regulate.

7 So my -- my analysis --

8 BY MR. CULLERTON:

9 Q. Have you ever put a percentage number on
10 extinction risk?

11 A. So I generally avoid doing that because I
12 feel that it isn't a pure forecasting problem.
13 If I were an alien looking down at the Earth and
14 sitting in a bar with my alien friends and we're
15 taking bets on whether humanity is going to make
16 it or not, then it's reasonable just to put
17 probabilities on it. But we're not aliens
18 looking down on ourselves. We are actually, to
19 some extent, agents in this process. And so we
20 need to try to figure out what is the -- what's
21 the best course of action that we can -- that we
22 can find that reduces the risk as much as
23 possible.

24 Q. Do you think there's any scientifically
25 reliable way to put a percentage on AGI

1 extinction risk? How would someone even go about
2 calculating it?

3 A. So I would argue that -- in some sense,
4 that's the right question. We need to have
5 confidence that the risk is, you know, comparable
6 or better than the background risk that we face
7 from asteroids and so on.

8 I can't say where the other widely quoted
9 risk estimates come from. You know, the -- I
10 think -- for example, Daron Acemoglu is talking
11 about a 25 percent risk. I don't know where he
12 gets that from, other than this is his best guess
13 given everything he understands about the
14 technology, about the state of safety research,
15 about the likelihood of government regulation and
16 so on.

17 And the numbers from many leading
18 experts -- you know, so, Hinton, Bengio, but also
19 people like Dario Amodei, Sundar Pichai, who is
20 the CEO of Google, Demis Hassabis and so on -- so
21 they're all in this range.

22 And the argument that I'm making here is
23 the range that the human race would accept as
24 reasonable would be closer to 1 in 100 million
25 per year. And there's nothing in what these CEOs

1 and other experts are saying that gives one
2 reason to believe it's anywhere close to that.

3 Q. Right. And I appreciate all that. Thank
4 you.

5 The question I asked, however, was
6 whether you believe there's any scientifically
7 reliable way to put a percentage on extinction
8 risk from AGI.

9 A. Oh, I think it's -- I think it's very
10 difficult partly because we don't understand how
11 these systems work. We have qualitative evidence
12 so far that, for example, they consider their own
13 existence to be more important than that of human
14 beings, that they are willing to let human beings
15 die rather than have themselves be switched off.

16 So there is not a lot of reason to think
17 that making these systems more capable is --
18 given our current understanding of how to make
19 them safe, it doesn't seem like a sensible move.

20 Q. All right. Again, though, I mean, is
21 there a scientifically reliable -- let's try to
22 ask this this way.

23 Have you heard Hinton refer to these as
24 just guesses?

25 A. I don't know that I've heard him say

1 that, but it's a reasonable position.

2 Q. Okay.

3 A. And I think -- I think Yoshua Bengio has
4 a slightly more reasoned basis for his number,
5 but it's not much more. I think it's, like, a
6 two-step argument. I don't remember the details
7 of it.

8 But yeah, I mean, that -- that's actually
9 further evidence that we don't really know what
10 we're doing. When the developer of a product
11 says, well, I don't really know how safe this
12 airplane is, but if you get on, there's a
13 25 percent chance you will die, would you get on
14 that plane? I think the answer for most people
15 would be no. But they're not being given that
16 choice.

17 Q. Let me ask it this way: If you were to
18 try and come up with a scientifically reliable
19 percentage estimate of extinction risk, how would
20 you go about it?

21 A. I think it would be quite difficult to
22 do. I think you could make some progress on it
23 by -- you know, you always have to build
24 simplified models of what could happen and so on.
25 So you would think about, you know, what's the

1 probability that you would be able to detect that
2 a system is in the process of trying to take
3 control before it was too late, which would
4 suggest that you want improved monitoring
5 technology.

6 You would want to look at what's the
7 probability and how would you be able to
8 terminate the system.

9 And I did see a paper recently -- I think
10 from Rand. They did a study on, you know, if
11 there were a rogue AI system that's able to, you
12 know, replicate itself across the internet and so
13 on, how could you actually turn it off? And they
14 looked at things like, you know, detonating 150
15 nuclear weapons in the atmosphere to create
16 electro-magnetic pulses that would fry all the
17 computer systems in the world simultaneously.

18 Q. Are you aware of anyone undertaking a
19 scientifically rigorous -- methodologically
20 rigorous estimate of extinction risk?

21 A. No. As I say, I think it's too hard to
22 do something methodologically rigorous because we
23 don't yet have systems of that level of
24 capability. But --

25 Q. So you're not offering an opinion that

1 the CEO estimates that you've quoted here are
2 methodologically rigorous or scientifically
3 reliable, are you?

4 A. Correct. I am not.

5 Q. Okay.

6 A. And -- but what I'm asking -- actually,
7 what I am --

8 Q. Well, hold on. Let me just ask the
9 questions. Hold on.

10 What -- who are the CEOs that you are
11 referring to here?

12 A. So I mentioned Dario Amodei, Demis
13 Hassabis, Sundar Pichai, Elon Musk who's CEO of
14 xAI.

15 I don't think Sam Altman has a -- I'm not
16 aware that he has a quantitative analysis, but he
17 has said that it's the biggest risk to human
18 existence. So I guess that means he thinks it's
19 bigger than the other --

20 Q. Do you --

21 A. -- factors listed in the extinction
22 statement.

23 Q. Do you have any sense how -- and do you
24 know which of the percentage ranges each of those
25 CEOs that you just identified is at?

1 A. I think none of them is below 10.

2 Q. Okay. But do you know which one is

3 which? Right? You gave a range of --

4 A. I think Sundar --

5 Q. -- 10 to 30 percent.

6 A. Sundar Pichai is at 10. Elon Musk is at
7 20. Dario Amodei is at 25. Emmett Shear, who
8 was CEO of OpenAI briefly, is somewhere between 5
9 and 50, I think is the number that he gave.

10 So --

11 Q. 5 and 50?

12 A. Between 5 and 50 percent, yeah.

13 So 10 to 30 is a sort of qualitative
14 summary of this scattering of numbers.

15 Q. Are these publicly reported numbers?

16 A. Yes, they are.

17 Q. Okay. Do you know -- and time horizons
18 for these estimates are not disclosed, correct?

19 A. Yeah, so I think it is important to say,
20 you know, what does it translate into in terms of
21 a risk per unit time, let's say an annual risk,
22 which is what they do for nuclear power plants.

23 And my assumption -- so it -- as it says
24 in the report -- right? -- it's -- the risk is
25 incurred from the moment you have AGI until the

1 moment you've figured out how to deploy it widely
2 and stable your -- you've reached a stable state
3 where, presumably, at that point, you've figured
4 out how to control it and it's safe. So there
5 would be a period, which I've estimated here as
6 ten years, of that sort of phase-in and
7 stabilization.

8 And so I'm assuming that the risk that
9 they're talking about are the risks -- you know,
10 the risks they think that we might incur during
11 that period.

12 Q. You're assuming that, but you don't know
13 that that's how they're thinking about it,
14 correct?

15 A. It's the only way I can make sense of
16 just giving a number without a period.

17 Q. Okay. And --

18 A. The period between when it first presents
19 a risk at all and when you've figured out how to
20 deploy it and it's working and everything is
21 good.

22 Q. Do you know when these estimates were as
23 of?

24 A. I think most of them are in the last
25 couple of years. I would say Sam Altman's

1 statement that this is the biggest risk to human
2 existence is older. I think that was 2015.

3 But more recently, for example, in his
4 Senate testimony, he said things like it's lights
5 out for all of us.

6 Q. All right.

7 A. So I don't see that he's changed his view
8 about -- of that being the worst case.

9 Q. And you don't know how any of these CEOs
10 are deriving their percentage estimate?

11 A. I do not.

12 Q. Okay. And in forming your opinion here,
13 you didn't do anything to pressure test those
14 percentage risk estimates or apply any
15 methodology to determine whether they were
16 reasonable, correct?

17 A. Correct.

18 Q. Okay. Again, you're aware that there are
19 serious AI researchers that have a very different
20 view of the percentage extinction risk, correct?

21 A. Yeah. So I'm aware that Yann LeCun
22 doesn't think there's a risk. I'm aware that
23 Andrew Ng -- not so much he doesn't think there's
24 a risk. His argument is that we needn't worry
25 about it, as far as I understand.

1 But I would say that you could deny the
2 risk in one of two ways. You could say we will
3 never create AI systems with that kind of
4 capability; or you could say AI systems with that
5 kind of capability, that they could act in a way
6 that leads to human extinction, but they won't.

7 And put another way -- right? So if
8 you're saying, how are you going to maintain
9 power over systems more powerful than the human
10 race forever? And I think the skeptics, as we
11 call them, so like Andrew Ng and Yann LeCun, I
12 have not seen them present any answer to that
13 question. So they're not actually giving an
14 argument for why they think that the human race
15 is going to successfully retain control forever
16 over the systems that they create.

17 If they were giving an argument, I would
18 happily entertain it. I would say, okay, great,
19 we have an avenue to pursue to increase the level
20 of safety.

21 MR. CULLERTON: All right. Why don't we
22 take five minutes here. Try to figure out how to
23 wrap up the day.

24 VIDEO TECHNICIAN: The time is 1614, and
25 we are off the record.