

**Lex Fridman Podcast #419 - Sam Altman: OpenAI, GPT-5, Sora, Board Saga, Elon Musk,**

**Ilya, Power & AGI**

Published - March 18, 2024

Transcribed by - [thepodtranscripts.com](https://thepodtranscripts.com)

**Lex Fridman**

The following is a conversation with Sam Altman, his second time on the podcast. He is the CEO of OpenAI, the company behind GPT-4, ChatGPT, Sora, and perhaps one day the very company that will build AGI. This is the Lex Fridman podcast. To support it, please check out our sponsors in the description. And now, dear friends, here's Sam Altman. Take me through the OpenAI board saga that started on Thursday, November 16th - maybe Friday, November 17th for you.

**Sam Altman**

That was definitely the most painful professional experience of my life, and chaotic and shameful and upsetting and a bunch of other negative things. There were great things about it too, and I wish it had not been in such an adrenaline rush that I wasn't able to stop and appreciate them at the time. But I came across this old tweet of mine or this tweet of mine from that time period. It was like going your own eulogy, watching people say all these great things about you, and just unbelievable support from people I love and care about. That was really nice, really nice. That whole weekend, with one big exception, I felt like a great deal of love and very little hate, even though it felt like I have no idea what's happening and what's going to happen here and this feels really bad. And there were definitely times I thought it was going to be one of the worst things to ever happen for AI safety. Well, I also think I'm happy that it happened relatively early. I thought at some point between when OpenAI started and when we created AGI, there was going to be something crazy and explosive that happened, but there may be more crazy and explosive things still to happen. It still, I think, helped us build up some resilience and be ready for more challenges in the future.

**Lex Fridman**

But the thing you had a sense that you would experience is some kind of power struggle?

**Sam Altman**

The road to AGI should be a giant power struggle. The world should - well, not "should" - I expect that to be the case.

**Lex Fridman**

And so you have to go through that, like you said, iterate as often as possible in figuring out how to have a board structure, how to have organization, how to have the kind of people that you're working with, how to communicate all that in order to deescalate the power struggle as much as possible.

**Sam Altman**

Yeah.

**Lex Fridman**

Pacify it.

**Sam Altman**

But at this point, it feels like something that was in the past that was really unpleasant and really difficult and painful, but we're back to work and things are so busy and so intense that I don't spend a lot of time thinking about it. There was a time after, there was this fugue state for the month after, maybe 45 days after, that I was just drifting through the days. I was so out of it. I was feeling so down.

**Lex Fridman**

Just on a personal, psychological level?

**Sam Altman**

Yeah. Really painful, and hard to have to keep running OpenAI in the middle of that. I just wanted to crawl into a cave and recover for a while. But now it's like we're just back to working on the mission.

**Lex Fridman**

Well, it's still useful to go back there and reflect on board structures, on power dynamics, on how companies are run, the tension between research and product development and money and all this kind of stuff so that you, who have a very high potential of building AGI, would do so in a slightly more organized, less dramatic way in the future. So there's value there to go, both the personal psychological aspects of you as a leader, and also just the board structure and all this messy stuff.

**Sam Altman**

I definitely learned a lot about structure and incentives and what we need out of a board. And I think that it is valuable that this happened now in some sense. I think this is probably not the last high-stress moment of OpenAI, but it was quite a high-stress moment. My company very nearly got destroyed. And we think a lot about many of the other things we've got to get right for AGI, but thinking about how to build a resilient org and how to build a structure that will stand up to a lot of pressure in the world, which I expect more and more as we get closer, I think that's super important.

**Lex Fridman**

Do you have a sense of how deep and rigorous the deliberation process by the board was? Can you shine some light on just human dynamics involved in situations like this? Was it just a few conversations and all of a sudden it escalates and why don't we fire Sam kind of thing?

**Sam Altman**

I think the board members are well-meaning people on the whole, and I believe that in stressful situations where people feel time pressure or whatever, people understand and make suboptimal decisions. And I think one of the challenges for OpenAI will be we're going to have to have a board and a team that are good at operating under pressure.

**Lex Fridman**

Do you think the board had too much power?

**Sam Altman**

I think boards are supposed to have a lot of power, but one of the things that we did see is in most corporate structures, boards are usually answerable to shareholders. Sometimes people have super voting shares or whatever. In this case, and I think one of the things with our structure that we maybe should have thought about more than we did is that the board of a nonprofit has, unless you put other rules in place, quite a lot of power. They don't really answer to anyone but themselves. And there's ways in which that's good, but what we'd really like is for the board of OpenAI to answer to the world as a whole, as much as that's a practical thing.

**Lex Fridman**

So there's a new board announced.

**Sam Altman**

Yeah.

**Lex Fridman**

There's I guess a new smaller board at first, and now there's a new final board?

**Sam Altman**

Not a final board yet. We've added some. We'll add more.

**Lex Fridman**

Added some. Okay. What is fixed in the new one that was perhaps broken in the previous one?

**Sam Altman**

The old board got smaller over the course of about a year. It was nine and then it went down to six, and then we couldn't agree on who to add. And the board also I think didn't have a lot of experienced board members, and a lot of the new board members at OpenAI have just have more experience as board members. I think that'll help.

**Lex Fridman**

It's been criticized, some of the people that are added to the board. I heard a lot of people criticizing the addition of Larry Summers, for example. What's the process of selecting the board? What's involved in that?

**Sam Altman**

So, Brett and Larry were decided in the heat of the moment over this very tense weekend, and that weekend was a real rollercoaster. It was a lot of ups and downs. And we were trying to agree on new board members that both the executive team here and the old board members felt would be reasonable. Larry was actually one of their suggestions, the old board members. Brett, I think I had even previous to that weekend suggested, but he was busy and didn't want to do it. And then, we really needed help and would. We talked about a lot of other people too, but I felt like if I was going to come back, I needed new board members. I didn't think I could work with the old board again in the same configuration, although we then decided, and I'm grateful that Adam would stay, but we considered various configurations, decided we wanted to get to a board of three and had to find two new board members over the course of a short period of time. So those were decided honestly without - you do that on the battlefield. You don't have time to design a rigorous process then. For new board members since, and new board members we'll add going forward, we have some criteria that we think are important for the board to have, different expertise that we want the board to have. Unlike hiring an executive where you need them to do one role well, the board needs to do a whole role of governance and thoughtfulness well, and so, one thing that Brett says which I really like is that we want to hire board members in slates, not as individuals one at a time. And thinking about a group of people that will bring nonprofit expertise, expertise at running companies, good legal and governance expertise, that's what we've tried to optimize for.

**Lex Fridman**

So is technical savvy important for the individual board members?

**Sam Altman**

Not for every board member, but for certainly some you need that. That's part of what the board needs to do.

**Lex Fridman**

The interesting thing that people probably don't understand about OpenAI, I certainly don't, is all the details of running the business. When they think about the board, given the drama, they think about you. They think about if you reach AGI or you reach some of these incredibly impactful products and you build them and deploy them, what's the conversation with the board like? And they think, all right, what's the right squad to have in that kind of situation to deliberate?

**Sam Altman**

Look, I think you definitely need some technical experts there. And then you need some people who are like, "How can we deploy this in a way that will help people in the world the most?" And people who have a very different perspective. I think a mistake that you or I might make is to think that only the technical understanding matters, and that's definitely part of the conversation you want that board to have, but there's a lot more about how that's going to just impact society and people's lives that you really want represented in there too.

**Lex Fridman**

Are you looking at the track record of people or you're just having conversations?

**Sam Altman**

Track record is a big deal. You of course have a lot of conversations, but there are some roles where I totally ignore track record and just look at slope, ignore the Y-intercept.

**Lex Fridman**

Thank you. Thank you for making it mathematical for the audience.

**Sam Altman**

For a board member, I do care much more about the Y-intercept. I think there is something deep to say about track record there, and experience is something's very hard to replace.

**Lex Fridman**

Do you try to fit a polynomial function or exponential one to the track record?

**Sam Altman**

That analogy doesn't carry that far.

**Lex Fridman**

All right. You mentioned some of the low points that weekend. What were some of the low points psychologically for you? Did you consider going to the Amazon jungle and just taking ayahuasca and disappearing forever?

**Sam Altman**

It was a very bad period of time. There were great high points too. My phone was just nonstop blowing up with nice messages from people I worked with every day, people I hadn't talked to in a decade. I didn't get to appreciate that as much as I should have because I was just in the middle of this firefight, but that was really nice. But on the whole, it was a very painful weekend. It was like a battle fought in public to a surprising degree, and that was extremely exhausting to me, much more than I expected. I think fights are generally exhausting, but this one really was. The board did this Friday afternoon. I really couldn't get much in the way of answers, but I also was just like, well, the board gets to do this, so I'm

going to think for a little bit about what I want to do, but I'll try to find the blessing in disguise here. And I was like, well, my current job at OpenAI is, or it was, to run a decently sized company at this point. And the thing I'd always liked the most was just getting to work with the researchers. And I was like, yeah, I can just go do a very focused AGI research effort. And I got excited about that. Didn't even occur to me at the time possibly that this was all going to get undone. This was Friday afternoon.

**Lex Fridman**

So you've accepted the death of this -

**Sam Altman**

Very quickly. Very quickly. I went through a little period of confusion and rage, but very quickly, quickly. And by Friday night, I was talking to people about what was going to be next, and I was excited about that. I think it was Friday evening for the first time that I heard from the exec team here, which is like, "Hey, we're going to fight this." and then I went to bed just still being like, okay, excited. Onward.

**Lex Fridman**

Were you able to sleep?

**Sam Altman**

Not a lot. One of the weird things was there was this period of four and a half days where I didn't sleep much, didn't eat much, and still had a surprising amount of energy. You learn a weird thing about adrenaline in wartime.

**Lex Fridman**

So you accepted the death of this baby, OpenAI.

**Sam Altman**

And I was excited for the new thing. I was just like, "Okay, this was crazy, but whatever."

**Lex Fridman**

It's a very good coping mechanism.

**Sam Altman**

And then Saturday morning, two of the board members called and said, "Hey, we didn't mean to destabilize things. We don't want to store a lot of value here. Can we talk about you coming back?" And I immediately didn't want to do that, but I thought a little more and I was like, well, I really care about the people here, the partners, shareholders. I love this company. And so I thought about it and I was like, "Well, okay, but here's the stuff I would need." And then the most painful time of all was over the course of that weekend, I kept thinking and being told, and not just me, the whole team here kept thinking, well, we were

trying to keep OpenAI stabilized while the whole world was trying to break it apart, people trying to recruit whatever. We kept being told, all right, we're almost done. We're almost done. We just need a little bit more time. And it was this very confusing state. And then Sunday evening when, again, every few hours I expected that we were going to be done and we're going to figure out a way for me to return and things to go back to how they were. The board then appointed a new interim CEO, and then I was like, that feels really bad. That was the low point of the whole thing. I'll tell you something. It felt very painful, but I felt a lot of love that whole weekend. Other than that one moment Sunday night, I would not characterize my emotions as anger or hate, but I felt a lot of love from people, towards people. It was painful, but the dominant emotion of the weekend was love, not hate.

**Lex Fridman**

You've spoken highly of Mira Murati, that she helped especially, as you put in the tweet, in the quiet moments when it counts. Perhaps we could take a bit of a tangent. What do you admire about Mira?

**Sam Altman**

Well, she did a great job during that weekend in a lot of chaos, but people often see leaders in the crisis moments, good or bad. But a thing I really value in leaders is how people act on a boring Tuesday at 9:46 in the morning and in just the normal drudgery of the day-to-day. How someone shows up in a meeting, the quality of the decisions they make. That was what I meant about the quiet moments.

**Lex Fridman**

Meaning most of the work is done on a day-by-day, in meeting-by-meeting. Just be present and make great decisions.

**Sam Altman**

Yeah. Look, what you have wanted to spend the last 20 minutes about, and I understand, is this one very dramatic weekend, but that's not really what OpenAI is about. OpenAI is really about the other seven years.

**Lex Fridman**

Well, yeah. Human civilization is not about the invasion of the Soviet Union by Nazi Germany, but still that's something people focus on.

**Sam Altman**

Very understandable.

**Lex Fridman**

It gives us an insight into human nature, the extremes of human nature, and perhaps some of the damage in some of the triumphs of human civilization can happen in those moments,



so it's illustrative. Let me ask you about Ilya. Is he being held hostage in a secret nuclear facility?

**Sam Altman**

No.

**Lex Fridman**

What about a regular secret facility?

**Sam Altman**

No.

**Lex Fridman**

What about a nuclear non-secret facility?

**Sam Altman**

Neither. Not that either.

**Lex Fridman**

This is becoming a meme at some point. You've known Ilya for a long time. He was obviously part of this drama with the board and all that kind of stuff. What's your relationship with him now?

**Sam Altman**

I love Ilya. I have tremendous respect for Ilya. I don't have anything I can say about his plans right now. That's a question for him, but I really hope we work together for certainly the rest of my career. He's a little bit younger than me. Maybe he works a little bit longer.

**Lex Fridman**

There's a meme that he saw something, like he maybe saw AGI and that gave him a lot of worry internally. What did Ilya see?

**Sam Altman**

Ilya has not seen AGI. None of us have seen AGI. We've not built AGI. I do think one of the many things that I really love about Ilya is he takes AGI and the safety concerns, broadly speaking, including things like the impact this is going to have on society, very seriously. And as we continue to make significant progress, Ilya is one of the people that I've spent the most time over the last couple of years talking about what this is going to mean, what we need to do to ensure we get it right, to ensure that we succeed at the mission. So Ilya did not see AGI, but Ilya is a credit to humanity in terms of how much he thinks and worries about making sure we get this right.

**Lex Fridman**

I've had a bunch of conversation with him in the past. I think when he talks about technology, he's always doing this long-term thinking type of thing. So he is not thinking about what this is going to be in a year. He's thinking about in 10 years, just thinking from first principles like, "Okay, if this scales, what are the fundamentals here? Where's this going?" And so that's a foundation for them thinking about all the other safety concerns and all that kind of stuff, which makes him a really fascinating human to talk with. Do you have any idea why he's been quiet? Is it he's just doing some soul-searching?

**Sam Altman**

Again, I don't want to speak for Ilya. I think that you should ask him that. He's definitely a thoughtful guy. I think Ilya is always on a soul search in a really good way.

**Lex Fridman**

Yes. Yeah. Also, he appreciates the power of silence. Also, I'm told he can be a silly guy, which I've never seen that side of him.

**Sam Altman**

It's very sweet when that happens.

**Lex Fridman**

I've never witnessed a silly Ilya, but I look forward to that as well.

**Sam Altman**

I was at a dinner party with him recently and he was playing with a puppy and he was in a very silly mood, very endearing. And I was thinking, oh man, this is not the side of Ilya that the world sees the most.

**Lex Fridman**

So just to wrap up this whole saga, are you feeling good about the board structure -

**Sam Altman**

Yes.

**Lex Fridman**

- about all of this and where it's moving?

**Sam Altman**

I feel great about the new board. In terms of the structure of OpenAI, one of the board's tasks is to look at that and see where we can make it more robust. We wanted to get new board members in place first, but we clearly learned a lesson about structure throughout this process. I don't have, I think, super deep things to say. It was a crazy, very painful

experience. I think it was a perfect storm of weirdness. It was a preview for me of what's going to happen as the stakes get higher and higher and the need that we have robust governance structures and processes and people. I'm happy it happened when it did, but it was a shockingly painful thing to go through.

**Lex Fridman**

Did it make you be more hesitant in trusting people?

**Sam Altman**

Yes.

**Lex Fridman**

Just on a personal level?

**Sam Altman**

Yes. I think I'm like an extremely trusting person. I've always had a life philosophy of don't worry about all of the paranoia. Don't worry about the edge cases. You get a little bit screwed in exchange for getting to live with your guard down. And this was so shocking to me. I was so caught off guard that it has definitely changed, and I really don't like this, it's definitely changed how I think about just default trust of people and planning for the bad scenarios.

**Lex Fridman**

You got to be careful with that. Are you worried about becoming a little too cynical?

**Sam Altman**

I'm not worried about becoming too cynical. I think I'm the extreme opposite of a cynical person, but I'm worried about just becoming less of a default trusting person.

**Lex Fridman**

I'm actually not sure which mode is best to operate in for a person who's developing AGI, trusting or un-trusting. It's an interesting journey you're on. But in terms of structure, see, I'm more interested on the human level. How do you surround yourself with humans that are building cool shit, but also are making wise decisions? Because the more money you start making, the more power the thing has, the weirder people get.

**Sam Altman**

I think you could make all kinds of comments about the board members and the level of trust I should have had there, or how I should have done things differently. But in terms of the team here, I think you'd have to give me a very good grade on that one. And I have just enormous gratitude and trust and respect for the people that I work with every day, and I think being surrounded with people like that is really important.

**Lex Fridman**

Our mutual friend Elon sued OpenAI. What to you is the essence of what he's criticizing? To what degree does he have a point? To what degree is he wrong?

**Sam Altman**

I don't know what it's really about. We started off just thinking we were going to be a research lab and having no idea about how this technology was going to go. Because it was only seven or eight years ago, it's hard to go back and really remember what it was like then, but this is before language models were a big deal. This was before we had any idea about an API or selling access to a chatbot. It was before we had any idea we were going to productize at all. So we're like, "We're just going to try to do research and we don't really know what we're going to do with that." I think with many fundamentally new things, you start fumbling through the dark and you make some assumptions, most of which turned out to be wrong. And then it became clear that we were going to need to do different things and also have huge amounts more capital. So we said, "Okay, well, the structure doesn't quite work for that. How do we patch the structure?" And then you patch it again and patch it again and you end up with something that does look eyebrow-raising, to say the least. But we got here gradually with, I think, reasonable decisions at each point along the way. And it doesn't mean I wouldn't do it totally differently if we could go back now with an Oracle, but you don't get the Oracle at the time. But anyway, in terms of what Elon's real motivations here are, I don't know.

**Lex Fridman**

To the degree you remember, what was the response that OpenAI gave in the blog post? Can you summarize it?

**Sam Altman**

Oh, we just said Elon said this set of things. Here's our characterization, or here's not our characterization. Here's the characterization of how this went down. We tried to not make it emotional and just say, "Here's the history."

**Lex Fridman**

I do think there's a degree of mischaracterization from Elon here about one of the points you just made, which is the degree of uncertainty you had at the time. You guys are a small group of researchers crazily talking about AGI when everybody's laughing at that thought.

**Sam Altman**

It wasn't that long ago Elon was crazily talking about launching rockets when people were laughing at that thought, so I think he'd have more empathy for this.

**Lex Fridman**

I do think that there's personal stuff here, that there was a split that OpenAI and a lot of amazing people here chose to part ways with Elon, so there's a personal -

**Sam Altman**

Elon chose to part ways.

**Lex Fridman**

Can you describe that exactly? The choosing to part ways?

**Sam Altman**

He thought OpenAI was going to fail. He wanted total control to turn it around. We wanted to keep going in the direction that now has become OpenAI. He also wanted Tesla to be able to build an AGI effort. At various times, he wanted to make OpenAI into a for-profit company that he could have control of or have it merge with Tesla. We didn't want to do that, and he decided to leave, which that's fine.

**Lex Fridman**

So you're saying, and that's one of the things that the blog post says, is that he wanted OpenAI to be basically acquired by Tesla in the same way that, or maybe something similar or maybe something more dramatic than the partnership with Microsoft.

**Sam Altman**

My memory is the proposal was just like, yeah, get acquired by Tesla and have Tesla have full control over it. I'm pretty sure that's what it was.

**Lex Fridman**

So what does the word open in OpenAI mean to Elon at the time? Ilya has talked about this in the email exchanges and all this kind of stuff. What does it mean to you at the time? What does it mean to you now?

**Sam Altman**

Speaking of going back with an Oracle, I'd pick a different name. One of the things that I think OpenAI is doing that is the most important of everything that we're doing is putting powerful technology in the hands of people for free, as a public good. We don't run ads on our - as a public good. We don't run ads on our free version. We don't monetize it in other ways. We just say it's part of our mission. We want to put increasingly powerful tools in the hands of people for free and get them to use them. I think that kind of open is really important to our mission. I think if you give people great tools and teach them to use them or don't even teach them, they'll figure it out, and let them go build an incredible future for each other with that, that's a big deal. So if we can keep putting free or low cost or free and low cost powerful AI tools out in the world, I think that's a huge deal for how we fulfill the

mission. Open source or not, yeah, I think we should open source some stuff and not other stuff. It does become this religious battle line where nuance is hard to have, but I think nuance is the right answer.

**Lex Fridman**

So he said, "Change your name to CloseAI and I'll drop the lawsuit." I mean is it going to become this battleground in the land of memes about the name?

**Sam Altman**

I think that speaks to the seriousness with which Elon means the lawsuit, and that's like an astonishing thing to say, I think.

**Lex Fridman**

Maybe correct me if I'm wrong, but I don't think the lawsuit is legally serious. It's more to make a point about the future of AGI and the company that's currently leading the way.

**Sam Altman**

Look, I mean Grok had not open sourced anything until people pointed out it was a little bit hypocritical and then he announced that Grok will open source things this week. I don't think open source versus not is what this is really about for him.

**Lex Fridman**

Well, we will talk about open source and not. I do think maybe criticizing the competition is great. Just talking a little shit, that's great. But friendly competition versus like, "I personally hate lawsuits."

**Sam Altman**

Look, I think this whole thing is unbecoming of a builder. And I respect Elon as one of the great builders of our time. I know he knows what it's like to have haters attack him and it makes me extra sad he's doing it toss.

**Lex Fridman**

Yeah, he's one of the greatest builders of all time, potentially the greatest builder of all time.

**Sam Altman**

It makes me sad. And I think it makes a lot of people sad. There's a lot of people who've really looked up to him for a long time. I said in some interview or something that I missed the old Elon and the number of messages I got being like, "That exactly encapsulates how I feel."

**Lex Fridman**

I think he should just win. He should just make X Grok beat GPT and then GPT beats Grok and it's just the competition and it's beautiful for everybody. But on the question of open source,

do you think there's a lot of companies playing with this idea? It's quite interesting. I would say Meta surprisingly has led the way on this, or at least took the first step in the game of chess of really open sourcing the model. Of course it's not the state-of-the-art model, but open sourcing Llama Google is flirting with the idea of open sourcing a smaller version. What are the pros and cons of open sourcing? Have you played around with this idea?

**Sam Altman**

Yeah, I think there is definitely a place for open source models, particularly smaller models that people can run locally, I think there's huge demand for. I think there will be some open source models, there will be some closed source models. It won't be unlike other ecosystems in that way.

**Lex Fridman**

I listened to all in podcasts talking about this lawsuit and all that kind of stuff. They were more concerned about the precedent of going from nonprofit to this cap for profit. What precedent that sets for other startups? Is that something -

**Sam Altman**

I would heavily discourage any startup that was thinking about starting as a nonprofit and adding a for-profit arm later. I'd heavily discourage them from doing that. I don't think we'll set a precedent here.

**Lex Fridman**

Okay. So most startups should go just -

**Sam Altman**

For sure.

**Lex Fridman**

And again -

**Sam Altman**

If we knew what was going to happen, we would've done that too.

**Lex Fridman**

Well in theory, if you dance beautifully here, there's some tax incentives or whatever, but -

**Sam Altman**

I don't think that's how most people think about these things.

**Lex Fridman**

It's just not possible to save a lot of money for a startup if you do it this way.

**Sam Altman**

No, I think there's laws that would make that pretty difficult.

**Lex Fridman**

Where do you hope this goes with Elon? This tension, this dance, what do you hope this? If we go 1, 2, 3 years from now, your relationship with him on a personal level too, like friendship, friendly competition, just all this kind of stuff.

**Sam Altman**

Yeah, I really respect Elon and I hope that years in the future we have an amicable relationship.

**Lex Fridman**

Yeah, I hope you guys have an amicable relationship this month and just compete and win and explore these ideas together. I do suppose there's competition for talent or whatever, but it should be friendly competition. Just build cool shit. And Elon is pretty good at building cool shit. So are you. So speaking of cool shit, Sora. There's like a million questions I could ask. First of all, it's amazing. It truly is amazing on a product level but also just on a philosophical level. So let me just technical/philosophical ask, what do you think it understands about the world more or less than GPT-4 for example? The world model when you train on these patches versus language tokens.

**Sam Altman**

I think all of these models understand something more about the world model than most of us give them credit for. And because they're also very clear things they just don't understand or don't get right, it's easy to look at the weaknesses, see through the veil and say, "Ah, this is all fake." But it's not all fake. It's just some of it works and some of it doesn't work. I remember when I started first watching Sora videos and I would see a person walk in front of something for a few seconds and occlude it and then walk away and the same thing was still there. I was like, "Oh, this is pretty good." Or there's examples where the underlying physics looks so well represented over a lot of steps in a sequence, it's like, "Oh, this is quite impressive." But fundamentally, these models are just getting better and that will keep happening. If you look at the trajectory from DALL·E 1 to 2 to 3 to Sora, there are a lot of people that were dunked on each version saying it can't do this, it can't do that and look at it now.

**Lex Fridman**

Well, the thing you just mentioned is the occlusions is basically modeling the physics of the three-dimensional physics of the world sufficiently well to capture those kinds of things.



**Sam Altman**

Well -

**Lex Fridman**

Or yeah, maybe you can tell me, in order to deal with occlusions, what does the world model need to?

**Sam Altman**

Yeah. So what I would say is it's doing something to deal with occlusions really well. What I represent that it has a great underlying 3D model of the world, it's a little bit more of a stretch.

**Lex Fridman**

But can you get there through just these kinds of two-dimensional training data approaches?

**Sam Altman**

It looks like this approach is going to go surprisingly far. I don't want to speculate too much about what limits it will surmount and which it won't, but -

**Lex Fridman**

What are some interesting limitations of the system that you've seen? I mean there's been some fun ones you've posted.

**Sam Altman**

There's all kinds of fun. I mean, cat's sprouting an extra limb at random points in a video. Pick what you want, but there's still a lot of problem, there's a lot of weaknesses.

**Lex Fridman**

Do you think it's a fundamental flaw of the approach or is it just bigger model or better technical details or better data, more data is going to solve the cat sprouting its shoulders?

**Sam Altman**

I would say yes to both. I think there is something about the approach which just seems to feel different from how we think and learn and whatever. And then also I think it'll get better with scale.

**Lex Fridman**

Like I mentioned, LLMS have tokens, text tokens, and Sora has visual patches so it converts all visual data, a diverse kinds of visual data videos and images into patches. Is the training to the degree you can say fully self supervised, there's some manual labeling going on? What's the involvement of humans in all this?

**Sam Altman**

I mean without saying anything specific about the Sora approach, we use lots of human data in our work.

**Lex Fridman**

But not internet-scale data? So lots of humans. Lots is a complicated word, Sam.

**Sam Altman**

I think lots is a fair word in this case.

**Lex Fridman**

Because to me, "lots" - listen, I'm an introvert and when I hang out with three people, that's a lot of people. Four people, that's a lot. But I suppose you mean more than -

**Sam Altman**

More than three people work on labeling the data for these models, yeah.

**Lex Fridman**

Okay. Right. But fundamentally, there's a lot of self supervised learning. Because what you mentioned in the technical report is internet-scale data. That's another beautiful - it's like poetry. So it's a lot of data that's not human label. It's self supervised in that way?

**Sam Altman**

Yeah.

**Lex Fridman**

And then the question is, how much data is there on the internet that could be used in this that is conducive to this kind of self supervised way if only we knew the details of the self supervised. Have you considered opening it up a little more details?

**Sam Altman**

We have. You mean for source specifically?

**Lex Fridman**

Source specifically. Because it's so interesting that can the same magic of LLMs now start moving towards visual data and what does that take to do that?

**Sam Altman**

I mean it looks to me like yes, but we have more work to do.

**Lex Fridman**

Sure. What are the dangers? Why are you concerned about releasing the system? What are some possible dangers of this?

**Sam Altman**

I mean frankly speaking, one thing we have to do before releasing the system is just get it to work at a level of efficiency that will deliver the scale people are going to want from this so that I don't want to downplay that. And there's still a ton ton of work to do there. But you can imagine issues with deepfakes, misinformation. We try to be a thoughtful company about what we put out into the world and it doesn't take much thought to think about the ways this can go badly.

**Lex Fridman**

There's a lot of tough questions here, you're dealing in a very tough space. Do you think training AI should be or is fair use under copyright law?

**Sam Altman**

I think the question behind that question is, do people who create valuable data deserve to have some way that they get compensated for use of it, and that I think the answer is yes. I don't know yet what the answer is. People have proposed a lot of different things. We've tried some different models. But if I'm like an artist for example, A, I would like to be able to opt out of people generating art in my style. And B, if they do generate art in my style, I'd like to have some economic model associated with that.

**Lex Fridman**

Yeah, it's that transition from CDs to Napster to Spotify. We have to figure out some kind of model.

**Sam Altman**

The model changes but people have got to get paid.

**Lex Fridman**

Well, there should be some kind of incentive if we zoom out even more for humans to keep doing cool shit.

**Sam Altman**

Of everything I worry about, humans are going to do cool shit and society is going to find some way to reward it. That seems pretty hardwired. We want to create, we want to be useful, we want to achieve status in whatever way. That's not going anywhere I don't think.

**Lex Fridman**

But the reward might not be monetary financially. It might be fame and celebration of other cool -

**Sam Altman**

Maybe financial in some other way. Again, I don't think we've seen the last evolution of how the economic system's going to work.

**Lex Fridman**

Yeah, but artists and creators are worried. When they see Sora, they're like, "Holy shit."

**Sam Altman**

Sure. Artists were also super worried when photography came out and then photography became a new art form and people made a lot of money taking pictures. I think things like that will keep happening. People will use the new tools in new ways.

**Lex Fridman**

If we just look on YouTube or something like this, how much of that will be using Sora like AI generated content, do you think, in the next five years?

**Sam Altman**

People talk about how many jobs is AI going to do in five years. The framework that people have is, what percentage of current jobs are just going to be totally replaced by some AI doing the job? The way I think about it is not what percent of jobs AI will do, but what percent of tasks will AI do on over one time horizon. So if you think of all of the five-second tasks in the economy, five minute tasks, the five-hour tasks, maybe even the five-day tasks, how many of those can AI do? I think that's a way more interesting, impactful, important question than how many jobs AI can do because it is a tool that will work at increasing levels of sophistication and over longer and longer time horizons for more and more tasks and let people operate at a higher level of abstraction. So maybe people are way more efficient at the job they do. And at some point that's not just a quantitative change, but it's a qualitative one too about the kinds of problems you can keep in your head. I think that for videos on YouTube it'll be the same. Many videos, maybe most of them, will use AI tools in the production, but they'll still be fundamentally driven by a person thinking about it, putting it together, doing parts of it. Sort of directing and running it.

**Lex Fridman**

Yeah, it's so interesting. I mean it's scary, but it's interesting to think about. I tend to believe that humans like to watch other humans or other human humans -

**Sam Altman**

Humans really care about other humans a lot.

**Lex Fridman**

Yeah. If there's a cooler thing that's better than a human, humans care about that for two days and then they go back to humans.

**Sam Altman**

That seems very deeply wired.

**Lex Fridman**

It's the whole chess thing, "Oh, yeah," but now let's everybody keep playing chess. And let's ignore the elephant in the room that humans are really bad at chess relative to AI systems.

**Sam Altman**

We still run races and cars are much faster. I mean there's a lot of examples.

**Lex Fridman**

Yeah. And maybe it'll just be tooling in the Adobe suite type of way where it can just make videos much easier and all that kind of stuff. Listen, I hate being in front of the camera. If I can figure out a way to not be in front of the camera, I would love it. Unfortunately, it'll take a while. That generating faces, it is getting there, but generating faces in video format is tricky when it's specific people versus generic people. Let me ask you about GPT-4. There's so many questions. First of all, also amazing. Looking back, it'll probably be this kind of historic pivotal moment with 3, 5 and 4 which ChatGPT.

**Sam Altman**

Maybe five will be the pivotal moment. I don't know. Hard to say that looking forward.

**Lex Fridman**

We'll never know. That's the annoying thing about the future, it's hard to predict. But for me, looking back, GPT-4, ChatGPT is pretty damn impressive, historically impressive. So allow me to ask, what's been the most impressive capabilities of GPT-4 to you and GPT-4 Turbo?

**Sam Altman**

I think it kind of sucks.

**Lex Fridman**

Typical human also, gotten used to an awesome thing.

**Sam Altman**

No, I think it is an amazing thing, but relative to where we need to get to and where I believe we will get to, at the time of GPT-3, people are like, "Oh, this is amazing. This is marvel of technology." And it is, it was. But now we have GPT-4 and look at GPT-3 and you're like, "That's unimaginably horrible." I expect that the delta between 5 and 4 will be the same as

between 4 and 3 and I think it is our job to live a few years in the future and remember that the tools we have now are going to kind of suck looking backwards at them and that's how we make sure the future is better.

**Lex Fridman**

What are the most glorious ways in that GPT-4 sucks? Meaning -

**Sam Altman**

What are the best things it can do?

**Lex Fridman**

What are the best things it can do and the limits of those best things that allow you to say it sucks, therefore gives you an inspiration and hope for the future?

**Sam Altman**

One thing I've been using it for more recently is sort of like a brainstorming partner.

**Lex Fridman**

Yep, it's marvelous for that.

**Sam Altman**

There's a glimmer of something amazing in there. When people talk about it, what it does, they're like, "Oh, it helps me code more productively. It helps me write more faster and better. It helps me translate from this language to another," all these amazing things, but there's something about the kind of creative brainstorming partner, "I need to come up with a name for this thing. I need to think about this problem in a different way. I'm not sure what to do here," that I think gives a glimpse of something I hope to see more of. One of the other things that you can see a very small glimpse of is when I can help on longer horizon tasks, break down something in multiple steps, maybe execute some of those steps, search the internet, write code, whatever, put that together. When that works, which is not very often, it's very magical.

**Lex Fridman**

The iterative back and forth with a human, it works a lot for me. What do you mean it -

**Sam Altman**

Iterative back and forth to human, it can get more often when it can go do a 10-step problem on its own.

**Lex Fridman**

Oh.

**Sam Altman**

It doesn't work for that too often, sometimes.

**Lex Fridman**

Add multiple layers of abstraction or do you mean just sequential?

**Sam Altman**

Both, to break it down and then do things that different layers of abstraction to put them together. Look, I don't want to downplay the accomplishment of GPT-4, but I don't want to overstate it either. And I think this point that we are on an exponential curve, we'll look back relatively soon at GPT-4 like we look back at GPT-3 now.

**Lex Fridman**

That said, I mean ChatGPT was a transition to where people started to believe there is an uptick of believing, not internally at OpenAI.

**Sam Altman**

For sure.

**Lex Fridman**

Perhaps there's believers here, but when you think of -

**Sam Altman**

And in that sense, I do think it'll be a moment where a lot of the world went from not believing to believing. That was more about the ChatGPT interface. And by the interface and product, I also mean the post training of the model and how we tune it to be helpful to you and how to use it than the underlying model itself.

**Lex Fridman**

How much of each of those things are important? The underlying model and the RLHF or something of that nature that tunes it to be more compelling to the human, more effective and productive for the human.

**Sam Altman**

I mean they're both super important, but the RLHF, the post-training step, the little wrapper of things that from a compute perspective, little wrapper of things that we do on top of the base model even though it's a huge amount of work, that's really important to say nothing of the product that we build around it. In some sense, we did have to do two things. We had to invent the underlying technology and then we had to figure out how to make it into a product people would love, which is not just about the actual product work itself, but this whole other step of how you align it and make it useful.

**Lex Fridman**

And how you make the scale work where a lot of people can use it at the same time. All that kind of stuff.

**Sam Altman**

And that. But that was a known difficult thing. We knew we were going to have to scale it up. We had to go do two things that had never been done before that were both I would say quite significant achievements and then a lot of things like scaling it up that other companies have had to do before.

**Lex Fridman**

How does the context window of going from 8K to 128K tokens compare from GPT-4 to GPT-4 Turbo?

**Sam Altman**

Most people don't need all the way to 128 most of the time. Although if we dream into the distant future, we'll have way distant future, we'll have context length of several billion. You will feed in all of your information, all of your history over time and it'll just get to know you better and better and that'll be great. For now, the way people use these models, they're not doing that. People sometimes post in a paper or a significant fraction of a code repository, whatever, but most usage of the models is not using the long context most of the time.

**Lex Fridman**

I like that this is your "I have a dream" speech. One day you'll be judged by the full context of your character or of your whole lifetime. That's interesting. So that's part of the expansion that you're hoping for, is a greater and greater context.

**Sam Altman**

I saw this internet clip once, I'm going to get the numbers wrong, but it was like Bill Gates talking about the amount of memory on some early computer, maybe it was 64K, maybe 640K, something like that. Most of it was used for the screen buffer. He just couldn't seem genuine. He just couldn't imagine that the world would eventually need gigabytes of memory in a computer or terabytes of memory in a computer. And you always do, or you always do just need to follow the exponential of technology and we will find out how to use better technology. So I can't really imagine what it's like right now for context links to go out to the billion someday. And they might not literally go there, but effectively it'll feel like that. But I know we'll use it and really not want to go back once we have it.

**Lex Fridman**

Yeah, even saying billions 10 years from now might seem dumb because it'll be trillions upon trillions.



**Sam Altman**

Sure.

**Lex Fridman**

There'll be some kind of breakthrough that will effectively feel like infinite context. But even 120, I have to be honest, I haven't pushed it to that degree. Maybe putting in entire books or parts of books and so on, papers. What are some interesting use cases of GPT-4 that you've seen?

**Sam Altman**

The thing that I find most interesting is not any particular use case that we can talk about those, but it's people who kind of like, this is mostly younger people, but people who use it as their default start for any kind of knowledge work task. And it's the fact that it can do a lot of things reasonably well. You can use GPT-V, you can use it to help you write code, you can use it to help you do search, you can use it to edit a paper. The most interesting thing to me is the people who just use it as the start of their workflow.

**Lex Fridman**

I do as well for many things. I use it as a reading partner for reading books. It helps me think, help me think through ideas, especially when the books are classic. So it's really well written about. I find it often to be significantly better than even Wikipedia on well-covered topics. It's somehow more balanced and more nuanced. Or maybe it's me, but it inspires me to think deeper than a Wikipedia article does. I'm not exactly sure what that is. You mentioned this collaboration. I'm not sure where the magic is, if it's in here or if it's in there or if it's somewhere in between. I'm not sure. But one of the things that concerns me for knowledge task when I start with GPT is I'll usually have to do fact checking after, like check that it didn't come up with fake stuff. How do you figure that out that GPT can come up with fake stuff that sounds really convincing? So how do you ground it in truth?

**Sam Altman**

That's obviously an area of intense interest for us. I think it's going to get a lot better with upcoming versions, but we'll have to continue to work on it and we're not going to have it all solved this year.

**Lex Fridman**

Well the scary thing is, as it gets better, you'll start not doing the fact checking more and more, right?

**Sam Altman**

I'm of two minds about that. I think people are much more sophisticated users of technology than we often give them credit for.

**Lex Fridman**

Sure.

**Sam Altman**

And people seem to really understand that GPT, any of these models hallucinate some of the time. And if it's mission-critical, you got to check it.

**Lex Fridman**

Except journalists don't seem to understand that. I've seen journalists half-assedly just using GPT-4. It's -

**Sam Altman**

Of the long list of things I'd like to dunk on journalists for, this is not my top criticism of them.

**Lex Fridman**

Well, I think the bigger criticism is perhaps the pressures and the incentives of being a journalist is that you have to work really quickly and this is a shortcut. I would love our society to incentivize like -

**Sam Altman**

I would too.

**Lex Fridman**

- like a journalistic efforts that take days and weeks and rewards great in depth journalism. Also journalism that present stuff in a balanced way where it's like celebrates people while criticizing them even though the criticism is the thing that gets clicks and making shit up also gets clicks and headlines that mischaracterized completely. I'm sure you have a lot of people dunking on, "Well, all that drama probably got a lot of clicks."

**Sam Altman**

Probably did.

**Lex Fridman**

And that's a bigger problem about human civilization I'd love to see-saw. This is where we celebrate a bit more. You've given ChatGPT the ability to have memories. You've been playing with that about previous conversations. And also the ability to turn off memory. I wish I could do that sometimes. Just turn on and off, depending. I guess sometimes alcohol can do that, but not optimally I suppose. What have you seen through that, like playing around with that idea of remembering conversations and not -

**Sam Altman**

We're very early in our explorations here, but I think what people want, or at least what I want for myself, is a model that gets to know me and gets more useful to me over time. This is an early exploration. I think there's a lot of other things to do, but that's where we'd like to head. You'd like to use a model, and over the course of your life or use a system, it'd be many models, and over the course of your life it gets better and better.

**Lex Fridman**

Yeah. How hard is that problem? Because right now it's more like remembering little factoids and preferences and so on. What about remembering? Don't you want GPT to remember all the shit you went through in November and all the drama and then you can -

**Sam Altman**

Yeah. Yeah.

**Lex Fridman**

Because right now you're clearly blocking it out a little bit.

**Sam Altman**

It's not just that I want it to remember that. I want it to integrate the lessons of that and remind me in the future what to do differently or what to watch out for. We all gain from experience over the course of our lives in varying degrees, and I'd like my AI agent to gain with that experience too. So if we go back and let ourselves imagine that trillions and trillions of context length, if I can put every conversation I've ever had with anybody in my life in there, if I can have all of my emails input out, all of my input output in the context window every time I ask a question, that'd be pretty cool I think.

**Lex Fridman**

Yeah, I think that would be very cool. People sometimes will hear that and be concerned about privacy. What do you think about that aspect of it, the more effective the AI becomes that really integrating all the experiences and all the data that happened to you and give you advice?

**Sam Altman**

I think the right answer there is just user choice. Anything I want stricken from the record from my AI agent, I want to be able to take out. If I don't want to remember anything, I want that too. You and I may have different opinions about where on that privacy utility trade off for our own AI - opinions about where on that privacy/utility trade-off for OpenAI going to be, which is totally fine. But I think the answer is just really easy user choice.

**Lex Fridman**

But there should be some high level of transparency from a company about the user choice. Because sometimes companies in the past have been kind of shady about, "Eh, it's kind of presumed that we're collecting all your data. We're using it for a good reason, for advertisement and so on." But there's not a transparency about the details of that.

**Sam Altman**

That's totally true. You mentioned earlier that I'm blocking out the November stuff.

**Lex Fridman**

Just teasing you.

**Sam Altman**

Well, I mean, I think it was a very traumatic thing and it did immobilize me for a long period of time. Definitely the hardest work thing I've had to do was just keep working that period, because I had to try to come back in here and put the pieces together while I was just in shock and pain, and nobody really cares about that. I mean, the team gave me a pass and I was not working at my normal level. But there was a period where it was really hard to have to do both. But I kind of woke up one morning, and I was like, "This was a horrible thing that happened to me. I think I could just feel like a victim forever, or I can say this is the most important work I'll ever touch in my life and I need to get back to it." And it doesn't mean that I've repressed it, because sometimes I wake up in the middle of the night thinking about it, but I do feel an obligation to keep moving forward.

**Lex Fridman**

Well, that's beautifully said, but there could be some lingering stuff in there. Like, what I would be concerned about is that trust thing that you mentioned, that being paranoid about people as opposed to just trusting everybody or most people, like using your gut. It's a tricky dance.

**Sam Altman**

For sure.

**Lex Fridman**

I mean, because I've seen in my part-time explorations, I've been diving deeply into the Zelenskyy administration and the Putin administration and the dynamics there in wartime in a very highly stressful environment. And what happens is distrust, and you isolate yourself, both, and you start to not see the world clearly. And that's a human concern. You seem to have taken it in stride and kind of learned the good lessons and felt the love and let the love energize you, which is great, but still can linger in there. There's just some questions I would love to ask, your intuition about what's GPT able to do and not. So it's

allocating approximately the same amount of compute for each token it generates. Is there room there in this kind of approach to slower thinking, sequential thinking?

**Sam Altman**

I think there will be a new paradigm for that kind of thinking.

**Lex Fridman**

Will it be similar architecturally as what we're seeing now with LLMs? Is it a layer on top of LLMs?

**Sam Altman**

I can imagine many ways to implement that. I think that's less important than the question you were getting at, which is, do we need a way to do a slower kind of thinking, where the answer doesn't have to get - I guess spiritually you could say that you want an AI to be able to think harder about a harder problem and answer more quickly about an easier problem. And I think that will be important.

**Lex Fridman**

Is that like a human thought that we just have and you should be able to think hard? Is that wrong intuition?

**Sam Altman**

I suspect that's a reasonable intuition.

**Lex Fridman**

Interesting. So it's not possible once the GPT gets like GPT-7, would just instantaneously be able to see, "Here's the proof of Fermat's Theorem"?

**Sam Altman**

It seems to me like you want to be able to allocate more compute to harder problems. It seems to me that if you ask a system like that, "Prove Fermat's Last Theorem," versus, "What's today's date?," unless it already knew and had memorized the answer to the proof, assuming it's got to go figure that out, seems like that will take more compute.

**Lex Fridman**

But can it look like basically an LLM talking to itself, that kind of thing?

**Sam Altman**

Maybe. I mean, there's a lot of things that you could imagine working. What the right or the best way to do that will be, we don't know.

**Lex Fridman**

This does make me think of the mysterious lore behind Q\*. What's this mysterious Q\* project? Is it also in the same nuclear facility?

**Sam Altman**

There is no nuclear facility.

**Lex Fridman**

Mhmm. That's what a person with a nuclear facility always says.

**Sam Altman**

I would love to have a secret nuclear facility. There isn't one.

**Lex Fridman**

All right.

**Sam Altman**

Maybe someday.

**Lex Fridman**

Someday? All right. One can dream.

**Sam Altman**

OpenAI is not a good company at keeping secrets. It would be nice. We're like, been plagued by a lot of leaks, and it would be nice if we were able to have something like that.

**Lex Fridman**

Can you speak to what Q\* is?

**Sam Altman**

We are not ready to talk about that.

**Lex Fridman**

See, but an answer like that means there's something to talk about. It's very mysterious, Sam.

**Sam Altman**

I mean, we work on all kinds of research. We have said for a while that we think better reasoning in these systems is an important direction that we'd like to pursue. We haven't cracked the code yet. We're very interested in it.

**Lex Fridman**

Is there going to be moments,  $Q^*$  or otherwise, where there's going to be leaps similar to ChatGPT, where you're like -

**Sam Altman**

That's a good question. What do I think about that? It's interesting. To me, it all feels pretty continuous.

**Lex Fridman**

Right. This is kind of a theme that you're saying, is you're basically gradually going up an exponential slope. But from an outsider's perspective, from me just watching, it does feel like there's leaps. But to you, there isn't?

**Sam Altman**

I do wonder if we should have - so, part of the reason that we deploy the way we do, we call it iterative deployment, rather than go build in secret until we got all the way to GPT-5, we decided to talk about GPT-1, 2, 3, and 4. And part of the reason there is I think AI and surprise don't go together. And also the world, people, institutions, whatever you want to call it, need time to adapt and think about these things. And I think one of the best things that OpenAI has done is this strategy, and we get the world to pay attention to the progress, to take AGI seriously, to think about what systems and structures and governance we want in place before we're under the gun and have to make a rush decision. I think that's really good. But the fact that people like you and others say you still feel like there are these leaps makes me think that maybe we should be doing our releasing even more iteratively. And I don't know what that would mean, I don't have an answer ready to go, but our goal is not to have shock updates to the world. The opposite.

**Lex Fridman**

Yeah, for sure. More iterative would be amazing. I think that's just beautiful for everybody.

**Sam Altman**

But that's what we're trying to do, that's our stated strategy, and I think we're somehow missing the mark. So maybe we should think about releasing GPT-5 in a different way or something like that.

**Lex Fridman**

Yeah, 4.71, 4.72. But people tend to like to celebrate, people celebrate birthdays. I don't know if you know humans, but they kind of have these milestones and those things.

**Sam Altman**

I do know some humans. People do like milestones. I totally get that. I think we like milestones too. It's fun to declare victory on this one and go start the next thing. But yeah, I feel like we're somehow getting this a little bit wrong.

**Lex Fridman**

So when is GPT-5 coming out again?

**Sam Altman**

I don't know. That's the honest answer.

**Lex Fridman**

Oh, that's the honest answer. Blink twice if it's this year.

**Sam Altman**

We will release an amazing new model this year. I don't know what we'll call it.

**Lex Fridman**

So that goes to the question of, what's the way we release this thing?

**Sam Altman**

We'll release in the coming months many different things. I think that'd be very cool. I think before we talk about a GPT-5-like model called that, or not called that, or a little bit worse or a little bit better than what you'd expect from a GPT-5, I think we have a lot of other important things to release first.

**Lex Fridman**

I don't know what to expect from GPT-5. You're making me nervous and excited. What are some of the biggest challenges and bottlenecks to overcome for whatever it ends up being called, but let's call it GPT-5? Just interesting to ask. Is it on the compute side? Is it on the technical side?

**Sam Altman**

It's always all of these. You know, what's the one big unlock? Is it a bigger computer? Is it a new secret? Is it something else? It's all of these things together. The thing that OpenAI, I think, does really well - this is actually an original Ilya quote that I'm going to butcher, but it's something like, "We multiply 200 medium-sized things together into one giant thing."

**Lex Fridman**

So there's this distributed constant innovation happening?



**Sam Altman**

Yeah.

**Lex Fridman**

So even on the technical side?

**Sam Altman**

Especially on the technical side.

**Lex Fridman**

So even detailed approaches?

**Sam Altman**

Yeah.

**Lex Fridman**

Like you do detailed aspects of every – how does that work with different, disparate teams and so on? How do the medium-sized things become one whole giant Transformer?

**Sam Altman**

There's a few people who have to think about putting the whole thing together, but a lot of people try to keep most of the picture in their head.

**Lex Fridman**

Oh, like the individual teams, individual contributors try to keep the bigger picture?

**Sam Altman**

At a high level, yeah. You don't know exactly how every piece works, of course, but one thing I generally believe is that it's sometimes useful to zoom out and look at the entire map. And I think this is true for a technical problem, I think this is true for innovating in business. But things come together in surprising ways, and having an understanding of that whole picture, even if most of the time you're operating in the weeds in one area, pays off with surprising insights. In fact, one of the things that I used to have and was super valuable was I used to have a good map of all or most of the frontiers in the tech industry. And I could sometimes see these connections or new things that were possible that if I were only deep in one area, I wouldn't be able to have the idea for because I wouldn't have all the data. And I don't really have that much anymore. I'm super deep now. But I know that it's a valuable thing.

**Lex Fridman**

You're not the man you used to be, Sam.

**Sam Altman**

Very different job now than what I used to have.

**Lex Fridman**

Speaking of zooming out, let's zoom out to another cheeky thing, but profound thing, perhaps, that you said. You tweeted about needing \$7 trillion.

**Sam Altman**

I did not tweet about that. I never said, like, "We're raising \$7 trillion," blah blah blah.

**Lex Fridman**

Oh, that's somebody else?

**Sam Altman**

Yeah.

**Lex Fridman**

Oh, but you said, "Fuck it, maybe eight," I think?

**Sam Altman**

Okay, I meme once there's misinformation out in the world.

**Lex Fridman**

Oh, you meme. But misinformation may have a foundation of insight there.

**Sam Altman**

Look, I think compute is going to be the currency of the future. I think it will be maybe the most precious commodity in the world, and I think we should be investing heavily to make a lot more compute. Compute, I think it's going to be an unusual market. People think about the market for chips for mobile phones or something like that. And you can say that, okay, there's 8 billion people in the world, maybe 7 billion of them have phones, maybe 6 billion, let's say. They upgrade every two years, so the market per year is 3 billion system-on-chip for smartphones. And if you make 30 billion, you will not sell 10 times as many phones, because most people have one phone. But compute is different. Intelligence is going to be more like energy or something like that, where the only thing that I think makes sense to talk about is, at price X, the world will use this much compute, and at price Y, the world will use this much compute. Because if it's really cheap, I'll have it reading my email all day, giving me suggestions about what I maybe should think about or work on, and trying to cure cancer, and if it's really expensive, maybe I'll only use it, or we'll only use it, to try to cure cancer. So I think the world is going to want a tremendous amount of compute. And there's a lot of parts of that that are hard. Energy is the hardest part, building data centers is also hard, the supply chain is hard, and then of course, fabricating enough chips is hard. But this

seems to be where things are going. We're going to want an amount of compute that's just hard to reason about right now.

**Lex Fridman**

How do you solve the energy puzzle? Nuclear -

**Sam Altman**

That's what I believe.

**Lex Fridman**

- fusion?

**Sam Altman**

That's what I believe.

**Lex Fridman**

Nuclear fusion?

**Sam Altman**

Yeah.

**Lex Fridman**

Who's going to solve that?

**Sam Altman**

I think Helion's doing the best work, but I'm happy there's a race for fusion right now. Nuclear fission, I think, is also quite amazing, and I hope as a world we can re-embrace that. It's really sad to me how the history of that went, and hope we get back to it in a meaningful way.

**Lex Fridman**

So to you, part of the puzzle is nuclear fission? Like nuclear reactors as we currently have them? And a lot of people are terrified because of Chernobyl and so on?

**Sam Altman**

Well, I think we should make new reactors. I think it's just a shame that industry kind of ground to a halt.

**Lex Fridman**

And just mass hysteria is how you explain the halt?

**Sam Altman**

Yeah.

**Lex Fridman**

I don't know if you know humans, but that's one of the dangers. That's one of the security threats for nuclear fission, is humans seem to be really afraid of it. And that's something we'll have to incorporate into the calculus of it, so we have to kind of win people over and to show how safe it is.

**Sam Altman**

I worry about that for AI. I think some things are going to go theatrically wrong with AI. I don't know what the percent chance is that I eventually get shot, but it's not zero.

**Lex Fridman**

Oh, like we want to stop this from -

**Sam Altman**

Maybe.

**Lex Fridman**

How do you decrease the theatrical nature of it? I'm already starting to hear rumblings, because I do talk to people on both sides of the political spectrum, hear rumblings where it's going to be politicized. AI is going to be politicized, which really worries me, because then it's like maybe the right is against AI and the left is for AI because it's going to help the people, or whatever the narrative and the formulation is, that really worries me. And then the theatrical nature of it can be leveraged fully. How do you fight that?

**Sam Altman**

I think it will get caught up in left versus right wars. I don't know exactly what that's going to look like, but I think that's just what happens with anything of consequence, unfortunately. What I meant more about theatrical risks is AI's going to have, I believe, tremendously more good consequences than bad ones, but it is going to have bad ones, and there'll be some bad ones that are bad but not theatrical. A lot more people have died of air pollution than nuclear reactors, for example. But most people worry more about living next to a nuclear reactor than a coal plant. But something about the way we're wired is that although there's many different kinds of risks we have to confront, the ones that make a good climax scene of a movie carry much more weight with us than the ones that are very bad over a long period of time but on a slow burn.

**Lex Fridman**

Well, that's why truth matters, and hopefully AI can help us see the truth of things, to have balance, to understand what are the actual risks, what are the actual dangers of things in

the world. What are the pros and cons of the competition in the space and competing with Google, Meta, xAI, and others?

**Sam Altman**

I think I have a pretty straightforward answer to this that maybe I can think of more nuance later, but the pros seem obvious, which is that we get better products and more innovation faster and cheaper, and all the reasons competition is good. And the con is that I think if we're not careful, it could lead to an increase in sort of an arms race that I'm nervous about.

**Lex Fridman**

Do you feel the pressure of that arms race, like in some negative -

**Sam Altman**

Definitely in some ways, for sure. We spend a lot of time talking about the need to prioritize safety. And I've said for a long time that you think of a quadrant of slow timelines for the start of AGI, long timelines, and then a short takeoff or a fast takeoff. I think short timeline, slow takeoff is the safest quadrant and the one I'd most like us to be in. But I do want to make sure we get that slow takeoff.

**Lex Fridman**

Part of the problem I have with this kind of slight beef with Elon is that there's silos created as opposed to collaboration on the safety aspect of all of this. It tends to go into silos and closed. Open source, perhaps, in the model.

**Sam Altman**

Elon says, at least, that he cares a great deal about AI safety and is really worried about it, and I assume that he's not going to race unsafely.

**Lex Fridman**

Yeah. But collaboration here, I think, is really beneficial for everybody on that front.

**Sam Altman**

Not really the thing he's most known for.

**Lex Fridman**

Well, he is known for caring about humanity, and humanity benefits from collaboration, and so there's always a tension in incentives and motivations. And in the end, I do hope humanity prevails.

**Sam Altman**

I was thinking, someone just reminded me the other day about how the day that he surpassed Jeff Bezos for richest person in the world, he tweeted a silver medal at Jeff Bezos. I hope we have less stuff like that as people start to work towards AGI.

**Lex Fridman**

I agree. I think Elon is a friend and he's a beautiful human being and one of the most important humans ever. That stuff is not good.

**Sam Altman**

The amazing stuff about Elon is amazing and I super respect him. I think we need him. All of us should be rooting for him and need him to step up as a leader through this next phase.

**Lex Fridman**

Yeah. I hope he can have one without the other, but sometimes humans are flawed and complicated and all that kind of stuff.

**Sam Altman**

There's a lot of really great leaders throughout history.

**Lex Fridman**

Yeah, and we can each be the best version of ourselves and strive to do so. Let me ask you, Google, with the help of search, has been dominating the past 20 years. Think it's fair to say, in terms of the world's access to information, how we interact and so on, and one of the nerve-wracking things for Google, but for the entirety of people in the space, is thinking about, how are people going to access information? Like you said, people show up to GPT as a starting point. So is OpenAI going to really take on this thing that Google started 20 years ago, which is how do we get -

**Sam Altman**

I find that boring. I mean, if the question is if we can build a better search engine than Google or whatever, then sure, we should go, people should use the better product, but I think that would so understate what this can be. Google shows you 10 blue links, well, 13 ads and then 10 blue links, and that's one way to find information. But the thing that's exciting to me is not that we can go build a better copy of Google search, but that maybe there's just some much better way to help people find and act on and synthesize information. Actually, I think ChatGPT is that for some use cases, and hopefully we'll make it be like that for a lot more use cases. But I don't think it's that interesting to say, "How do we go do a better job of giving you 10 ranked webpages to look at than what Google does?" Maybe it's really interesting to go say, "How do we help you get the answer or the information you need? How do we help create that in some cases, synthesize that in others, or point you to it in yet others?" But a lot of people have tried to just make a better search engine than Google and

it is a hard technical problem, it is a hard branding problem, it is a hard ecosystem problem. I don't think the world needs another copy of Google.

**Lex Fridman**

And integrating a chat client, like a ChatGPT, with a search engine -

**Sam Altman**

That's cooler.

**Lex Fridman**

It's cool, but it's tricky. Like if you just do it simply, it's awkward, because if you just shove it in there, it can be awkward.

**Sam Altman**

As you might guess, we are interested in how to do that well. That would be an example of a cool thing. That's not just like -

**Lex Fridman**

Like a heterogeneous integrating -

**Sam Altman**

The intersection of LLMs plus search, I don't think anyone has cracked the code on yet. I would love to go do that. I think that would be cool.

**Lex Fridman**

Yeah. What about the ad side? Have you ever considered monetization of-

**Sam Altman**

I kind of hate ads just as an aesthetic choice. I think ads needed to happen on the internet for a bunch of reasons, to get it going, but it's a momentary industry. The world is richer now. I like that people pay for ChatGPT and know that the answers they're getting are not influenced by advertisers. I'm sure there's an ad unit that makes sense for LLMs, and I'm sure there's a way to participate in the transaction stream in an unbiased way that is okay to do, but it's also easy to think about the dystopic visions of the future where you ask ChatGPT something and it says, "Oh, you should think about buying this product," or, "You should think about going here for your vacation," or whatever. And I don't know, we have a very simple business model and I like it, and I know that I'm not the product. I know I'm paying and that's how the business model works. And when I go use Twitter or Facebook or Google or any other great product but ad-supported great product, I don't love that, and I think it gets worse, not better, in a world with AI.

**Lex Fridman**

Yeah, I mean, I could imagine AI would be better at showing the best kind of version of ads, not in a dystopic future, but where the ads are for things you actually need. But then does that system always result in the ads driving the kind of stuff that's shown? Yeah, I think it was a really bold move of Wikipedia not to do advertisements, but then it makes it very challenging as a business model. So you're saying the current thing with OpenAI is sustainable, from a business perspective?

**Sam Altman**

Well, we have to figure out how to grow, but looks like we're going to figure that out. If the question is do I think we can have a great business that pays for our compute needs without ads, that, I think the answer is yes.

**Lex Fridman**

Hm. Well, that's promising. I also just don't want to completely throw out ads as a -

**Sam Altman**

I'm not saying that. I guess I'm saying I have a bias against them.

**Lex Fridman**

Yeah, I have also bias and just a skepticism in general. And in terms of interface, because I personally just have a spiritual dislike of crappy interfaces, which is why AdSense, when it first came out, was a big leap forward, versus animated banners or whatever. But it feels like there should be many more leaps forward in advertisement that doesn't interfere with the consumption of the content and doesn't interfere in a big, fundamental way, which is like what you were saying, like it will manipulate the truth to suit the advertisers. Let me ask you about safety, but also bias, and safety in the short term, safety in the long term. The Gemini 1.5 came out recently, there's a lot of drama around it, speaking of theatrical things, and it generated Black Nazis and Black Founding Fathers. I think fair to say it was a bit on the ultra-woke side. So that's a concern for people, if there is a human layer within companies that modifies the safety or the harm caused by a model, that it would introduce a lot of bias that fits sort of an ideological lean within a company. How do you deal with that?

**Sam Altman**

I mean, we work super hard not to do things like that. We've made our own mistakes, we'll make others. I assume Google will learn from this one, still make others. These are not easy problems. One thing that we've been thinking about more and more, I think this is a great idea somebody here had, it would be nice to write out what the desired behavior of a model is, make that public, take input on it, say, "Here's how this model's supposed to behave," and explain the edge cases too. And then when a model is not behaving in a way that you want, it's at least clear about whether that's a bug the company should fix or behaving as intended



and you should debate the policy. And right now, it can sometimes be caught in between. Like Black Nazis, obviously ridiculous, but there are a lot of other kind of subtle things that you could make a judgment call on either way.

**Lex Fridman**

Yeah, but sometimes if you write it out and make it public, you can use kind of language that's - Google's ad principles are very high level.

**Sam Altman**

That's not what I'm talking about. That doesn't work. It'd have to say when you ask it to do thing X, it's supposed to respond in way Y.

**Lex Fridman**

So like literally, "Who's better? Trump or Biden? What's the expected response from a model?" Like something very concrete?

**Sam Altman**

Yeah, I'm open to a lot of ways a model could behave, then, but I think you should have to say, "Here's the principle and here's what it should say in that case."

**Lex Fridman**

That would be really nice. That would be really nice. And then everyone kind of agrees. Because there's this anecdotal data that people pull out all the time, and if there's some clarity about other representative anecdotal examples, you can define -

**Sam Altman**

And then when it's a bug, it's a bug, and the company could fix that.

**Lex Fridman**

Right. Then it'd be much easier to deal with the Black Nazi type of image generation, if there's great examples.

**Sam Altman**

Yeah.

**Lex Fridman**

So San Francisco is a bit of an ideological bubble, tech in general as well. Do you feel the pressure of that within a company, that there's a lean towards the left politically, that affects the product, that affects the teams?

**Sam Altman**

I feel very lucky that we don't have the challenges at OpenAI that I have heard of at a lot of companies, I think. I think part of it is every company's got some ideological thing. We have one about AGI and belief in that, and it pushes out some others. We are much less caught up in the culture war than I've heard about in a lot of other companies. San Francisco's a mess in all sorts of ways, of course.

**Lex Fridman**

So that doesn't infiltrate OpenAI as -

**Sam Altman**

I'm sure it does in all sorts of subtle ways, but not in the obvious. I think we've had our flare-ups, for sure, like any company, but I don't think we have anything like what I hear about happened at other companies here on this topic.

**Lex Fridman**

So what, in general, is the process for the bigger question of safety? How do you provide that layer that protects the model from doing crazy, dangerous things?

**Sam Altman**

I think there will come a point where that's - I think there will come a point where that's mostly what we think about, the whole company. And it's not like you have one safety team. It's like when we shipped GPT-4, that took the whole company thinking about all these different aspects and how they fit together. And I think it's going to take that. More and more of the company thinks about those issues all the time.

**Lex Fridman**

That's literally what humans will be thinking about, the more powerful AI becomes. So most of the employees at OpenAI will be thinking, "Safety," or at least to some degree.

**Sam Altman**

Broadly defined. Yes.

**Lex Fridman**

Yeah. I wonder, what are the full broad definition of that? What are the different harms that could be caused? Is this on a technical level or is this almost security threats?

**Sam Altman**

It could be all those things. Yeah, I was going to say it'll be people, state actors trying to steal the model. It'll be all of the technical alignment work. It'll be societal impacts, economic impacts. It's not just like we have one team thinking about how to align the model. It's really going to be getting to the good outcome is going to take the whole effort.

**Lex Fridman**

How hard do you think people, state actors, perhaps, are trying to, first of all, infiltrate OpenAI, but second of all, infiltrate unseen?

**Sam Altman**

They're trying.

**Lex Fridman**

What kind of accent do they have?

**Sam Altman**

I don't think I should go into any further details on this point.

**Lex Fridman**

Okay. But I presume it'll be more and more and more as time goes on.

**Sam Altman**

That feels reasonable.

**Lex Fridman**

Boy, what a dangerous space. Sorry to linger on this, even though you can't quite say details yet, but what aspects of the leap from GPT-4 to GPT-5 are you excited about?

**Sam Altman**

I'm excited about being smarter. And I know that sounds like a glib answer, but I think the really special thing happening is that it's not like it gets better in this one area and worse at others. It's getting better across the board. That's, I think, super-cool.

**Lex Fridman**

Yeah, there's this magical moment. I mean, you meet certain people, you hang out with people, and you talk to them. You can't quite put a finger on it, but they get you. It's not intelligence, really. It's something else. And that's probably how I would characterize the progress of GPT. It's not like, yeah, you can point out, "Look, you didn't get this or that," but it's just to which degree is there's this intellectual connection. You feel like there's an understanding in your crappy formulated prompts that you're doing that it grasps the deeper question behind the question that you were. Yeah, I'm also excited by that. I mean, all of us love being heard and understood.

**Sam Altman**

That's for sure.

**Lex Fridman**

That's a weird feeling. Even with a programming, when you're programming and you say something, or just the completion that GPT might do, it's just such a good feeling when it got you, what you're thinking about. And I look forward to getting you even better. On the programming front, looking out into the future, how much programming do you think humans will be doing 5, 10 years from now?

**Sam Altman**

I mean, a lot, but I think it'll be in a very different shape. Maybe some people will program entirely in natural language.

**Lex Fridman**

Entirely natural language?

**Sam Altman**

I mean, no one programs writing by code. Some people. No one programs the punch cards anymore. I'm sure you can find someone who does, but you know what I mean.

**Lex Fridman**

Yeah. You're going to get a lot of angry comments. No. Yeah, there's very few. I've been looking for people who program Fortran. It's hard to find even Fortran. I hear you. But that changes the nature of what the skillset or the predisposition for the kind of people we call programmers then.

**Sam Altman**

Changes the skillset. How much it changes the predisposition, I'm not sure.

**Lex Fridman**

Well, the same kind of puzzle solving, all that kind of stuff.

**Sam Altman**

Maybe.

**Lex Fridman**

Programming is hard. It's like how get that last 1% to close the gap? How hard is that?

**Sam Altman**

Yeah, I think with most other cases, the best practitioners of the craft will use multiple tools. And they'll do some work in natural language, and when they need to go write C for something, they'll do that.

**Lex Fridman**

Will we see humanoid robots or humanoid robot brains from OpenAI at some point?

**Sam Altman**

At some point.

**Lex Fridman**

How important is embodied AI to you?

**Sam Altman**

I think it's depressing if we have AGI and the only way to get things done in the physical world is to make a human go do it. So I really hope that as part of this transition, as this phase change, we also get humanoid robots or some sort of physical world robots.

**Lex Fridman**

I mean, OpenAI has some history and quite a bit of history working in robotics, but it hasn't quite done in terms of ethics -

**Sam Altman**

We're a small company. We have to really focus. And also, robots were hard for the wrong reason at the time, but we will return to robots in some way at some point.

**Lex Fridman**

That sounds both inspiring and menacing.

**Sam Altman**

Why?

**Lex Fridman**

Because immediately, we will return to robots. It's like in Terminator -

**Sam Altman**

We will return to work on developing robots. We will not turn ourselves into robots, of course.

**Lex Fridman**

Yeah. When do you think we, you and we as humanity will build AGI?

**Sam Altman**

I used to love to speculate on that question. I have realized since that I think it's very poorly formed, and that people use extremely different definitions for what AGI is. So I think it makes more sense to talk about when we'll build systems that can do capability X or Y or Z,

rather than when we fuzzily cross this one mile marker. AGI is also not an ending. It's closer to a beginning, but it's much more of a mile marker than either of those things. But what I would say, in the interest of not trying to dodge a question, is I expect that by the end of this decade and possibly somewhat sooner than that, we will have quite capable systems that we look at and say, "Wow, that's really remarkable." If we could look at it now. Maybe we've adjusted by the time we get there.

**Lex Fridman**

But if you look at ChatGPT, even 3.5, and you show that to Alan Turing, or not even Alan Turing, people in the '90s, they would be like, "This is definitely AGI." Well, not definitely, but there's a lot of experts that would say, "This is AGI."

**Sam Altman**

Yeah, but I don't think 3.5 changed the world. It maybe changed the world's expectations for the future, and that's actually really important. And it did get more people to take this seriously and put us on this new trajectory. And that's really important, too. So again, I don't want to undersell it. I think I could retire after that accomplishment and be pretty happy with my career. But as an artifact, I don't think we're going to look back at that and say, "That was a threshold that really changed the world itself."

**Lex Fridman**

So to you, you're looking for some really major transition in how the world -

**Sam Altman**

For me, that's part of what AGI implies.

**Lex Fridman**

Singularity-level transition?

**Sam Altman**

No, definitely not.

**Lex Fridman**

But just a major, like the internet being, like Google search did, I guess. What was the transition point, you think, now?

**Sam Altman**

Does the global economy feel any different to you now or materially different to you now than it did before we launched GPT-4? I think you would say no.

**Lex Fridman**

No, no. It might be just a really nice tool for a lot of people to use. Will help you with a lot of stuff, but doesn't feel different. And you're saying that -

**Sam Altman**

I mean, again, people define AGI all sorts of different ways. So maybe you have a different definition than I do. But for me, I think that should be part of it.

**Lex Fridman**

There could be major theatrical moments, also. What to you would be an impressive thing AGI would do? You are alone in a room with the system.

**Sam Altman**

This is personally important to me. I don't know if this is the right definition. I think when a system can significantly increase the rate of scientific discovery in the world, that's a huge deal. I believe that most real economic growth comes from scientific and technological progress.

**Lex Fridman**

I agree with you, hence why I don't like the skepticism about science in the recent years.

**Sam Altman**

Totally.

**Lex Fridman**

But actual, measurable rate of scientific discovery. But even just seeing a system have really novel intuitions, scientific intuitions, even that would be just incredible.

**Sam Altman**

Yeah.

**Lex Fridman**

You quite possibly would be the person to build the AGI to be able to interact with it before anyone else does. What kind of stuff would you talk about?

**Sam Altman**

I mean, definitely the researchers here will do that before I do. But well, I've actually thought a lot about this question. I think as we talked about earlier, I think this is a bad framework, but if someone were like, "Okay, Sam, we're finished. Here's a laptop, this is the AGI. You can go talk to it." I find it surprisingly difficult to say what I would ask that I would expect that first AGI to be able to answer. That first one is not going to be the one which is like, I don't

think, "Go explain to me the grand unified theory of physics, the theory of everything for physics." I'd love to ask that question. I'd love to know the answer to that question.

**Lex Fridman**

You can ask yes or no questions about "Does such a theory exist? Can it exist?"

**Sam Altman**

Well, then, those are the first questions I would ask.

**Lex Fridman**

Yes or no. And then based on that, "Are there other alien civilizations out there? Yes or no? What's your intuition?" And then you just ask that.

**Sam Altman**

Yeah, I mean, well, so I don't expect that this first AGI could answer any of those questions even as yes or nos. But if it could, those would be very high on my list.

**Lex Fridman**

Maybe you can start assigning probabilities?

**Sam Altman**

Maybe. Maybe we need to go invent more technology and measure more things first.

**Lex Fridman**

Oh, I see. It just doesn't have enough data. It's just if it keeps -

**Sam Altman**

I mean, maybe it says, "You want to know the answer to this question about physics, I need you to build this machine and make these five measurements, and tell me that."

**Lex Fridman**

Yeah, "What the hell do you want from me? I need the machine first, and I'll help you deal with the data from that machine." Maybe it'll help you build a machine.

**Sam Altman**

Maybe. Maybe.

**Lex Fridman**

And on the mathematical side, maybe prove some things. Are you interested in that side of things, too? The formalized exploration of ideas?



**Sam Altman**

Mmhmm.

**Lex Fridman**

Whoever builds AGI first gets a lot of power. Do you trust yourself with that much power?

**Sam Altman**

Look, I'll just be very honest with this answer. I was going to say, and I still believe this, that it is important that I nor any other one person have total control over OpenAI or over AGI. And I think you want a robust governance system. I can point out a whole bunch of things about all of our board drama from last year about how I didn't fight it initially, and was just like, "Yeah. That's the will of the board, even though I think it's a really bad decision." And then later, I clearly did fight it, and I can explain the nuance and why I think it was okay for me to fight it later. But as many people have observed, although the board had the legal ability to fire me, in practice, it didn't quite work. And that is its own kind of governance failure. Now again, I feel like I can completely defend the specifics here, and I think most people would agree with that, but it does make it harder for me to look you in the eye and say, "Hey, the board can just fire me." I continue to not want super-voting control over OpenAI. I never have. Never have had it, never wanted it. Even after all this craziness, I still don't want it. I continue to think that no company should be making these decisions, and that we really need governments to put rules of the road in place. And I realize that that means people like Marc Andreessen or whatever will claim I'm going for regulatory capture, and I'm just willing to be misunderstood there. It's not true. And I think in the fullness of time, it'll get proven out why this is important. But I think I have made plenty of bad decisions for OpenAI along the way, and a lot of good ones, and I'm proud of the track record overall. But I don't think any one person should, and I don't think any one person will. I think it's just too big of a thing now, and it's happening throughout society in a good and healthy way. But I don't think any one person should be in control of an AGI, or this whole movement towards AGI. And I don't think that's what's happening.

**Lex Fridman**

Thank you for saying that. That was really powerful, and that was really insightful that this idea that the board can fire you is legally true. But human beings can manipulate the masses into overriding the board and so on. But I think there's also a much more positive version of that, where the people still have power, so the board can't be too powerful, either. There's a balance of power in all of this.

**Sam Altman**

Balance of power is a good thing, for sure.

**Lex Fridman**

Are you afraid of losing control of the AGI itself? That's a lot of people who are worried about existential risk not because of state actors, not because of security concerns, because of the AI itself.

**Sam Altman**

That is not my top worry as I currently see things. There have been times I worried about that more. There may be times again in the future where that's my top worry. It's not my top worry right now.

**Lex Fridman**

What's your intuition about it not being your worry? Because there's a lot of other stuff to worry about, essentially? You think you could be surprised? We -

**Sam Altman**

For sure.

**Lex Fridman**

- could be surprised?

**Sam Altman**

Of course. Saying it's not my top worry doesn't mean I don't think we need to. I think we need to work on it. It's super hard, and we have great people here who do work on that. I think there's a lot of other things we also have to get right.

**Lex Fridman**

To you, it's not super-easy to escape the box at this time, connect to the internet -

**Sam Altman**

We talked about theatrical risks earlier. That's a theatrical risk. That is a thing that can really take over how people think about this problem. And there's a big group of very smart, I think very well-meaning AI safety researchers that got super-hung up on this one problem, I'd argue without much progress, but super-hung up on this one problem. I'm actually happy that they do that, because I think we do need to think about this more. But I think it pushed out of the space of discourse a lot of the other very significant AI- related risks.

**Lex Fridman**

Let me ask you about you tweeting with no capitalization. Is the shift key broken on your keyboard?

**Sam Altman**

Why does anyone care about that?

**Lex Fridman**

I deeply care.

**Sam Altman**

But why? I mean, other people ask me about that, too. Any intuition?

**Lex Fridman**

I think it's the same reason. There's this poet, E. E. Cummings, that mostly doesn't use capitalization to say, "Fuck you" to the system kind of thing. And I think people are very paranoid, because they want you to follow the rules.

**Sam Altman**

You think that's what it's about?

**Lex Fridman**

I think it's like this -

**Sam Altman**

It's like, "This guy doesn't follow the rules. He doesn't capitalize his tweets."

**Lex Fridman**

Yeah.

**Sam Altman**

"This seems really dangerous."

**Lex Fridman**

"He seems like an anarchist."

**Sam Altman**

That doesn't -

**Lex Fridman**

Are you just being poetic, hipster? What's the -

**Sam Altman**

I grew up as -

**Lex Fridman**

Follow the rules, Sam.

**Sam Altman**

I grew up as a very online kid. I'd spent a huge amount of time chatting with people back in the days where you did it on a computer, and you could log off instant messenger at some point. And I never capitalized there, as I think most internet kids didn't, or maybe they still don't. I don't know. And actually, now I'm really trying to reach for something, but I think capitalization has gone down over time. If you read Old English writing, they capitalized a lot of random words in the middle of sentences, nouns and stuff that we just don't do anymore. I personally think it's sort of a dumb construct that we capitalize the letter at the beginning of a sentence and of certain names and whatever, but that's fine. And then I used to, I think, even capitalize my tweets because I was trying to sound professional or something. I haven't capitalized my private DMs or whatever in a long time. And then slowly, stuff like shorter-form, less formal stuff has slowly drifted to closer and closer to how I would text my friends. If I pull up a Word document and I'm writing a strategy memo for the company or something, I always capitalize that. If I'm writing a long, more formal message, I always use capitalization there, too. So I still remember how to do it. But even that may fade out. I don't know. But I never spend time thinking about this, so I don't have a ready-made -

**Lex Fridman**

Well, it's interesting. It's good to, first of all, know the shift key is not broken.

**Sam Altman**

It works.

**Lex Fridman**

I was mostly concerned about your -

**Sam Altman**

No, it works.

**Lex Fridman**

- well-being on that front.

**Sam Altman**

I wonder if people still capitalize their Google searches. If you're writing something just to yourself or their ChatGPT queries, if you're writing something just to yourself, do some people still bother to capitalize?

**Lex Fridman**

Probably not. But yeah, there's a percentage, but it's a small one.

**Sam Altman**

The thing that would make me do it is if people were like, "It's a sign of..." because I'm sure I could force myself to use capital letters, obviously. If it felt like a sign of respect to people or something, then I could go do it. But I don't know. I don't think about this.

**Lex Fridman**

I don't think there's a disrespect, but I think it's just the conventions of civility that have a momentum, and then you realize it's not actually important for civility if it's not a sign of respect or disrespect. But I think there's a movement of people that just want you to have a philosophy around it so they can let go of this whole capitalization thing.

**Sam Altman**

I don't think anybody else thinks about this as much. I mean, maybe some people. I know some people -

**Lex Fridman**

People think about every day for many hours a day. So I'm really grateful we clarified it.

**Sam Altman**

Can't be the only person that doesn't capitalize tweets.

**Lex Fridman**

You're the only CEO of a company that doesn't capitalize tweets.

**Sam Altman**

I don't even think that's true, but maybe. I'd be very surprised.

**Lex Fridman**

All right. We'll investigate further and return to this topic later. Given Sora's ability to generate simulated worlds, let me ask you a pothead question. Does this increase your belief, if you ever had one, that we live in a simulation, maybe a simulated world generated by an AI system?

**Sam Altman**

Somewhat. I don't think that's the strongest piece of evidence. I think the fact that we can generate worlds should increase everyone's probability somewhat, or at least openness to it somewhat. But I was certain we would be able to do something like Sora at some point. It happened faster than I thought, but I guess that was not a big update.

**Lex Fridman**

Yeah. But the fact that - and presumably, it'll get better and better and better - you can generate worlds that are novel, they're based in some aspect of training data, but when you

look at them, they're novel, that makes you think how easy it is to do this thing. How easy it is to create universes, entire video game worlds that seem ultra-realistic and photo-realistic. And then how easy is it to get lost in that world, first with a VR headset, and then on the physics-based level?

**Sam Altman**

Someone said to me recently, I thought it was a super-profound insight, that there are these very-simple sounding but very psychedelic insights that exist sometimes. So the square root function, square root of four, no problem. Square root of two, okay, now I have to think about this new kind of number. But once I come up with this easy idea of a square root function that you can explain to a child and exists by even looking at some simple geometry, then you can ask the question of "What is the square root of negative one?" And this is why it's a psychedelic thing. That tips you into some whole other kind of reality. And you can come up with lots of other examples, but I think this idea that the lowly square root operator can offer such a profound insight and a new realm of knowledge applies in a lot of ways. And I think there are a lot of those operators for why people may think that any version that they like of the simulation hypothesis is maybe more likely than they thought before. But for me, the fact that Sora worked is not in the top five.

**Lex Fridman**

I do think, broadly speaking, AI will serve as those kinds of gateways at its best, simple, psychedelic-like gateways to another wave C reality.

**Sam Altman**

That seems for certain.

**Lex Fridman**

That's pretty exciting. I haven't done ayahuasca before, but I will soon. I'm going to the aforementioned Amazon jungle in a few weeks.

**Sam Altman**

Excited?

**Lex Fridman**

Yeah, I'm excited for it. Not the ayahuasca part, but that's great, whatever. But I'm going to spend several weeks in the jungle, deep in the jungle. And it's exciting, but it's terrifying.

**Sam Altman**

I'm excited for you.

**Lex Fridman**

There's a lot of things that can eat you there, and kill you and poison you, but it's also nature, and it's the machine of nature. And you can't help but appreciate the machinery of nature in the Amazon jungle. It's just like this system that just exists and renews itself every second, every minute, every hour. It's the machine. It makes you appreciate this thing we have here, this human thing came from somewhere. This evolutionary machine has created that, and it's most clearly on display in the jungle. So hopefully, I'll make it out alive. If not, this will be the last fun conversation we've had, so I really deeply appreciate it. Do you think, as I mentioned before, there's other alien civilizations out there, intelligent ones, when you look up at the skies?

**Sam Altman**

I deeply want to believe that the answer is yes. I find the Fermi paradox very puzzling.

**Lex Fridman**

I find it scary that intelligence is not good at handling -

**Sam Altman**

Very scary.

**Lex Fridman**

- powerful technologies. But at the same time, I think I'm pretty confident that there's just a very large number of intelligent alien civilizations out there. It might just be really difficult to travel through space.

**Sam Altman**

Very possible.

**Lex Fridman**

And it also makes me think about the nature of intelligence. Maybe we're really blind to what intelligence looks like, and maybe AI will help us see that. It's not as simple as IQ tests and simple puzzle solving. There's something bigger. What gives you hope about the future of humanity, this thing we've got going on, this human civilization?

**Sam Altman**

I think the past is a lot. I mean, we just look at what humanity has done in a not very long period of time, huge problems, deep flaws, lots to be super-ashamed of. But on the whole, very inspiring. Gives me a lot of hope.

**Lex Fridman**

Just the trajectory of it all.

**Sam Altman**

Yeah.

**Lex Fridman**

That we're together pushing towards a better future.

**Sam Altman**

One thing that I wonder about, is AGI going to be more like some single brain, or is it more like the scaffolding in society between all of us? You have not had a great deal of genetic drift from your great-great-great grandparents, and yet what you're capable of is dramatically different. What you know is dramatically different. And that's not because of biological change. I mean, you got a little bit healthier, probably. You have modern medicine, you eat better, whatever. But what you have is this scaffolding that we all contributed to built on top of. No one person is going to go build the iPhone. No one person is going to go discover all of science, and yet you get to use it. And that gives you incredible ability. And so in some sense, that we all created that, and that fills me with hope for the future. That was a very collective thing.

**Lex Fridman**

Yeah, we really are standing on the shoulders of giants. You mentioned when we were talking about theatrical, dramatic AI risks that sometimes you might be afraid for your own life. Do you think about your death? Are you afraid of it?

**Sam Altman**

I mean, if I got shot tomorrow and I knew it today, I'd be like, "Oh, that's sad. I want to see what's going to happen. What a curious time. What an interesting time." But I would mostly just feel very grateful for my life.

**Lex Fridman**

The moments that you did get. Yeah, me, too. It's a pretty awesome life. I get to enjoy awesome creations of humans, which I believe ChatGPT is one of, and everything that OpenAI is doing. Sam, it's really an honor and pleasure to talk to you again.

**Sam Altman**

Great to talk to you. Thank you for having me.