

[Transcript] Lex Fridman Podcast / #383 - Mark Zuckerberg: Future of AI at Meta, Facebook, Instagram, and WhatsApp

The following is a conversation with Mark Zuckerberg, his second time in this podcast.

He's the CEO of Meta that owns Facebook, Instagram, and WhatsApp, all services used by billions of people to connect with each other.

We talk about his vision for the future of Meta and the future of AI in our human world.

And now a quick few second mention of each sponsor. Check them out in the description.

It's the best way to support this podcast. We got Numeri for the world's hardest data science tournament, Shopify for e-commerce, and BetterHelp for mental health. Choose wisely, my friends.

Also, if you want to work with our amazing team, we're always hiring, go to lexfreedman.com

slash hiring. And now onto the full ad reads. As always, no ads in the middle. I find those annoying.

But these here ads, I try to make interesting. Though you may skip them if you must, my friends,

but please still check out the sponsors. They help this podcast out. I enjoy their stuff. Maybe you

will too. This show is brought to you by Numeri, a hedge fund that uses AI and machine learning

to make investment decisions. I'm a huge fan of real world data sets and real world machine

learning competitions to figure out what works. This is not ImageNet. This is not an artificial

toy data set for the development of toy systems that illustrate toy concepts. Those are the early,

early, early stages of research. But when you really want to see what works, you want benchmarks

that have stakes, that have the highest of stakes, especially ones that have money involved. So

I'm a huge fan, money or not, of data sets that represent the real world and demonstrate that

the system can operate in the real world at the highest of stakes. That's why I was really interested

in autonomous vehicles when the stakes are life and death with safety critical systems,

incredibly exciting to work on systems that are truly real world data sets. Anyway,

if that kind of thing interests you, if you're a machine learning engineer, head over to

Numeri.ai slash Lex to sign up for a tournament and hone your machine learning skills. That's

N-U-M-E-R dot AI slash Lex for a chance to play against me and win share of the tournament prize pool.

This show is also brought to you by Shopify, a platform designed for anyone to sell anywhere

with a great looking online store that brings your ideas to life and tools to manage the day-to-day

operations. Operations is such a badass word. I feel like you're running things. Anyway,

a few folks asked me about merch. I'm a huge fan of buying merch for the podcasts, shows, bands I

love and so I love the camaraderie of merch and I think Shopify is a great place to sell merch.

I'm definitely going to put out some merch. I'm really sorry I've been taking forever.

I've been working with this incredible artist. I just love art. I love artistic representation

of the funny, the profound on the t-shirt that allows you to celebrate with others. Something

super cool. I love it. To me, there's nothing like promotional about it, all that kind of stuff.

It's just sharing your happiness. Anyway, so I'll definitely use Shopify to create a merch store

so that people can share a bit of their happiness with others. If you have stuff to sell or you

have merch to sell or you want to share some of your happiness with others, sign up for a one

dollar per month trial period at Shopify.com slash Lex. That's all lowercase. Go to Shopify.com

slash Lex to take your business to the next level. This episode is also brought to you by Better Help,

spelled H-E-L-P Help. They figure out what you need to match you with a licensed professional

therapist in under 48 hours. I do a podcast. Obviously, I'm a big fan of talk therapy. In fact,

when I just listen to podcasts, it's a kind of talk therapy because I'm having a conversation with

the people I'm listening to in my mind. Whenever it's an interview shown as two folks talking,

I'm always the third person in the room, kind of almost participating in the conversation. And there's something therapeutic about that. So if you're listening to two other people tell their life stories and you be able to project your trauma, your struggles, your hopes, your dreams, your triumphs, all that kind of stuff onto their life and kind of dance with that. Of course, to do that rigorously and really just put it all out there in a raw and honest way, I think that's what therapy is about. There's a lot of things you can do for your mental health, but therapy is one of the obvious things you should have in the toolkit of lifestyle flourishing. Anyway, Better Help just makes the whole thing super easy. Super easy to sign up, super easy to find, a licensed therapist, all of that. It's obviously discreet, it's easy, it's affordable, it's available anywhere. Check them out at [betterhelp.com slash lex](https://betterhelp.com/lex) and save when you first month. That's [betterhelp.com slash lex](https://betterhelp.com/lex). This is the Lex Friedman podcast. And now, dear friends, here's Mark Zuckerberg. So, you competed in your first jujitsu tournament and me as a fellow jujitsu practitioner and competitor, I think that's really inspiring given all the things you have going on. So, I got to ask, what was that experience like? Oh, it was fun. I mean, I'm a pretty competitive person. Doing sports that basically require your full attention, I think is really important to my mental health and the way I just stay focused at doing everything I'm doing. So, I decided to get into martial arts and it's awesome. I got a ton of my friends into it. We all trained together. We have a mini academy in my garage. And I guess one of my friends was like, hey, we should go do a tournament. I was like, okay, yeah, let's do it. I'm not gonna shy away from a challenge like that. So, yeah, it was awesome. It was just a lot of fun. You weren't scared? There was no fear? I don't know. I was pretty sure that I'd do okay. I like the confidence. Well, so for people who don't know, jujitsu is a martial art where you're trying to break your opponent's limbs or choke them to sleep and do so with grace and elegance and efficiency and all that kind of stuff. It's a kind of art form, I think. They can do for your whole life and it's basically a game, a sport of human chess. You can think of there's a lot of strategy. There's a lot of sort of interesting human dynamics of using leverage and all that kind of stuff. And it's kind of incredible what you could do. You could do things like a small opponent could defeat a much larger opponent and you get to understand the way the mechanics of the human body works because of that. But you certainly can't be distracted. No. It's 100% focused sport. To compete, I needed to get around the fact that I didn't want it to be this big thing. So, I basically just, I rolled up with a hat and sunglasses and I was wearing a COVID mask and I registered under my first and middle name, so Mark Elliott. And it wasn't until I actually pulled all that stuff off right before I got on the mat that I think people knew it was me. So, it was pretty low key. But you're still a public figure. Yeah. I mean, I didn't want to lose. Right. The thing you're partially afraid of is not just the losing but being almost like embarrassed. It's so raw, the sport, in that like it's just you and another human being. There's a primal aspect there. Oh, yeah. It's great.

For a lot of people, it can be terrifying, especially the first time you're doing the common competing and you wasn't for you. I see the look of excitement in your face.

Yeah, I don't know. No fear.

I just think part of learning is failing.

So, I mean, the main thing, people who train jiu-jitsu, it's like you need to not have pride because all the stuff that you were talking about before about getting choked or getting a joint lock, it's you only get into a bad situation if you're not willing to tap once you've already lost. But obviously, when you're getting started with something, you're not going to be an expert at it immediately. So, you just need to be willing to go with that. But I think this is like, I don't know. I mean, maybe I've just been embarrassed enough times in my life.

I do think that there's a thing where like, as people grow up, maybe they don't want to be embarrassed or anything. They've built their adult identity and they kind of have a sense of who they are and what they want to project. And I don't know, I think maybe to some degree, you know, your ability to keep doing interesting things is your willingness to be embarrassed again and go back to step one and start as a beginner and get your ass kicked and look stupid doing things. And I think so many of the things that we're doing, whether it's this, I mean, this is just like a kind of a physical part of my life, but running the company, it's like we just take on new adventures. And all the big things that we're doing, I think of as like 10 plus year missions that we're on where, you know, often early on, you know, people doubt that we're going to be able to do it. And the initial work seems kind of silly. And our whole ethos says we don't want to wait until something is perfect to put it out there. We want to get it out quickly and get feedback on it. And so I don't know. I mean, there's probably just something about how I approach things in there. But I just kind of think that the moment that you decide that you're going to be too embarrassed to try something new,

then you're not going to learn anything anymore. But like I mentioned, that fear, that anxiety could be there, it could creep up every once in a while. Do you feel that in, especially stressful moments sort of outside of, did you just, Matt, just at work, stressful moments, big decision days, big decision moments? How do you deal with that fear? How do you deal with that anxiety? The thing that stresses me out the most is always the people challenges. You know, I kind of think that, you know, strategy questions, you know, I tend to have enough conviction around the values of what we're trying to do and what I think matters and what I want our company to stand for, that those don't really keep me up at night that much. I mean, I kind of, you know, it's not that I get everything right. Of course I don't, right? I mean, we make a lot of mistakes. But I at least have a pretty strong sense of where I want us to go on that. The thing, and running a company for, you know, almost 20 years now, one of the things that's been pretty clear is when you have a team that's cohesive, you can get almost anything done. And, you know, you can, you can run through super hard challenges. You can make hard decisions and push really hard to do the best work even, you know, and kind of optimize something super well. But when, when there's that tension, I mean, that's, that's when, when things get really tough. And, you know, when I talk to other friends who run other companies and things like that, I think one of the things that I actually spend a disproportionate amount of time on and running this company is just fostering a

pretty tight core group of people who are running the company with me. And that to me is, is kind of the thing that both makes it fun, right? Having, having, you know, friends and people you've worked with for a while and new people and new perspectives, but like a pretty tight group who

can, who you can go work on some of these crazy things with. But to me, that's also the most stressful thing is when, when there, when there's tension, you know, that's, that, that weighs on me. I think the, you know, just it's, it's maybe not surprising. I mean, we're like a very people focused company. And it's the, the people is the, the part of it that, that, you know, weighs on me the most to make sure that we get right. But yeah, that, that, that I'd say across everything that we do is probably the, the big thing. So when there's tension in that inner circle of, of close folks. So when you trust those folks to help you make difficult decisions about Facebook, WhatsApp, Instagram, the future of the company and the metaverse with AI, how do you build that close-knit group of folks to make those difficult decisions? Is there people that you have to have critical voices, very different perspectives on focusing on the past versus the future, all that kind of stuff? Yeah. I mean, I think for one thing, it's just spending a lot of time with whatever the group is that you want to be that core group, grappling with all of the biggest challenges. And that requires a fair amount of openness. And, you know, so I mean, a lot of how I run the company is, you know, it's like every Monday morning we get our, it's about the top 30 people together. And we, and this is a group that just worked together for a long period of time. And I mean, people, people rotate in, I mean, we knew people join, people leave the company, people go do other roles in the company. So it's not the same group over time. But then we spend, you know, a lot of times a couple of hours, a lot of the time it's, you know, it can be somewhat unstructured. We, like, I'll come with maybe a few topics that I, that are top of mind for me, but I'll ask other people to bring things and people, you know, raise questions, whether it's okay, there's an issue happening in some country with some policy issue. There's like a new technology that's developing here. We're having an issue with this partner. You know, there's a design trade-off and WhatsApp between two things that, that end up being values that we care about deeply. And we need to kind of decide where we want to be on that. And I just think over time, when, you know, by working through a lot of issues with people and doing it openly, people develop an intuition for each other and a bond and camaraderie. And to me, developing that is, is like a lot of the fun part of running a company or doing anything, right? I think it's like having, having people who are kind of along on the journey that you're, that you feel like you're doing it with, nothing is ever just one person doing it. Other people that disagree often within that group. It's a fairly combative group. Okay. So combat is part of it. So this is making decisions on design, engineering, policy, everything. Yeah, everything, everything. Yeah. I have to ask just back to Jiu-Jitsu for a little bit. What's your favorite submission? Now that you've been doing it, what's, how do you like to submit your opponent, Mark Zuckerberg? I mean, what, but first of all, do you prefer no Gi or Gi Jiu-Jitsu? So Gi is this outfit you wear that is maybe mimics clothing so you can choke. It's like a kimono. It's like the traditional martial arts or kimono. Yeah, pajamas. Pajamas. That you could choke people with, yes. Well, it's got the lapels. Yes. Yeah. So I like Jiu-Jitsu. I also really like MMA. And so I think no Gi more closely approximates MMA. And I think my style is, is maybe a little closer to an MMA

style. So like a lot of Jiu-Jitsu players are fine being on their back, right? And obviously, having a good guard is a critical part of Jiu-Jitsu. But in MMA, you don't want to be on your back, right? Because even if you have control, you're just taking punches while you're on your back. So that's no good. Do you like being on top? My style is I'm probably more pressure. And yeah, and I'd probably rather be the top player. But I'm also smaller, right? I'm not like a heavyweight guy, right? So from that perspective, I think, like, you know, it's especially because, you know, if I'm doing a competition, I'll compete with people who are my size, but you know, a lot of my friends are bigger than me. So back takes probably pretty important, right? Because that's where you have the most leverage advantage, right? Where people, you know, their arms, your arms are very weak behind you, right? So being able to get to the back and take that pretty important. But I don't know, I feel like the right strategy is to not be too committed to any single submission. But that said, I don't like hurting people. So I always think that chokes are a somewhat more humane way to go than joint locks. Yeah, and it's more about control. It's less dynamic. So you're basically like a Habib Narmangamatov type of fighter. So let's go, yeah, backtake to a rear naked choke. I think it's like the clean way to go. Straight forward answer right there. What advice would you give to people looking to start learning Jiu Jitsu? Given how busy you are, given where you are in life, you're able to do this, you're able to train, you're able to compete and get to learn something from this interesting art. Why do you think you have to be willing to just get beaten up a lot? Yeah. But I mean, over time, I think that there's a flow to all these things. And there's, you know, one of my experiences that I think kind of transcends running a company and the different activities that I like doing are, I really believe that if you're going to accomplish whatever anything, a lot of it is just being willing to push through and having the grit and determination to push through difficult situations. I think for a lot of people that that ends up being sort of a difference maker between the people who kind of get the most done and not. I mean, there's all these questions about like, you know, how many days people want to work and things like that. I think almost all the people who like start successful companies or things like that are just are working extremely hard. But I think one of the things that you learn both by doing this over time or, you know, very acutely with things like Jiu Jitsu or surfing is you can't push through everything. And I think that you learn this stuff very acutely doing sports compared to running a company because running a company, the cycle times are so long where it's like you start a project and then, you know, it's like months later or, you know, if you're building hardware, it could be years later before you're actually getting feedback and able to make the next set of decisions for the next version of the thing that you're doing. Whereas you're one of the things that I just think is mentally so nice about these very high turnaround conditioning, sports, things like that is you get feedback very quickly, right? It's like, okay, like I don't counter something correctly, you get punched in the face, right? So not in Jiu Jitsu, you don't get punched in Jiu Jitsu, but in MMA. There are all these analogies between all these things that I think actually hold that are that are like important life lessons, right? It's like, okay, you're surfing a wave. It's like, you know, sometimes you're like, you can't go in the other direction on it, right? It's like, there are limits to kind of what, you know, it's like a foil. You can pump the foil

and push pretty hard in a bunch of directions, but like, yeah, you, you know, it's at some level, like the momentum against you is strong enough, you're, that's not going to work. And I do think that that's sort of a humbling, but also an important lesson for, and I think people who are running things or building things, it's like, yeah, you, you know, a lot of the game is just being able to kind of push and work through complicated things, but you also need to kind of have enough of an understanding of like, which things you just can't push through and where the finesse is more important. Yeah. What are your Jiu Jitsu life lessons?

Well, I think you did it. You made it sound so simple and we're so eloquent that it's easy to miss, but basically being okay and accepting the wisdom and the joy in the getting your ass kicked in the full range of what that means. I think that's a big gift of the being humbled. Somehow being humbled, especially physically opens your mind to the full process of learning, what it means to learn, which is being willing to suck at something. I think Jiu Jitsu just very repetitively, efficiently humbles you over and over and over and over to where you can carry that lessons to places where you don't get humbled as much, whether it's research or running a company or building stuff, the cycle is longer. And Jiu Jitsu, you can just get humbled in a period of an hour over and over and over and over, especially when you're a beginner, you'll have a little person, just somebody much smarter than you, just kick your ass repeatedly, definitively, where there's no argument. Oh yeah. And then you literally tap because if you don't tap, you're going to die. So this is an agreement. You could have killed me just now, but we're friends. So we're going to agree that you're not going to. And that kind of humbling process, it just does something to your psyche, to your ego that puts it in its proper context to realize that everything in this life is like a journey from sucking through a hard process of improving rigorously day after day after day after day. Any kind of success requires hard work.

Yeah, Jiu Jitsu, more than a lot of sports, I would say, because I've done a lot of them, really teaches you that. And you made it sound so simple. It's okay, it's part of the process, you just get humbled, get your ass kicked. I've just failed and been embarrassed so many times in my life that it's a core competence of this. It's a core competence. Well, yes. And there's a deep truth to that, being able to, and you said it in the very beginning, which is, that's the thing that stops us, especially as you get older, especially as you develop expertise in certain areas, not being willing to be a beginner in a new area. Because that's where the growth happens, is being willing to be a beginner, being willing to be embarrassed, saying something stupid, doing something stupid. A lot of us that get good at one thing, you want to show that off. And it sucks being a beginner, but it's where growth happens.

Well, speaking of which, let me ask you about AI. It seems like this year, for the entirety of the human civilization, is an interesting year for the development of artificial intelligence.

A lot of interesting stuff is happening. So Meta is a big part of that. Meta has developed Lama, which is a 65 billion parameter model. There's a lot of interesting questions that can ask here, one of which has to do with open source. But first, can you tell the story of developing of this model and making the complicated decision of how to release it?

Yeah, sure. I think you're right, first of all, that in the last year, there have been a bunch of advances on scaling up these large transformer models. So there's the language equivalent of it with large language models. There's sort of the image generation equivalent with these large diffusion models. There's a lot of fundamental research that's gone into this. And Meta has

taken the approach of being quite open and academic in our development of AI. Part of this is we want to have the best people in the world researching this. And a lot of the best people want to know that they're going to be able to share their work. So that's part of the deal that we have, is that if you're one of the top AI researchers in the world and come here, you can get access to industry scale infrastructure. And part of our ethos is that we want to share what's invented broadly. We do that with a lot of the different AI tools that we create. And Lama is the language model that our research team made. And we did a limited open source release for it, which was intended for researchers to be able to use it. But responsibility and getting safety right on these is very important. So we didn't think that for the first one, there were a bunch of questions around whether we should be releasing this commercially. So we kind of punted on that for V1 of Lama and just released it from research. Now, obviously, by releasing it for research, it's out there. But companies know that they're not supposed to kind of put it into commercial releases. And we're working on the follow-up models for this and thinking through how exactly this should work for follow-up. Now that we've had time to work on a lot more of the safety and the pieces around that. But overall, I mean, this is, I just kind of think that it would be good if there were a lot of different folks who had the ability to build state-of-the-art technology here. And not just a small number of big companies. Where to train one of these AI models, the state-of-the-art models, is, just takes hundreds of millions of dollars of infrastructure. So there are not that many organizations in the world that can do that at the biggest scale today. And now it gets more efficient every day. So I do think that that will be available to more folks over time. But I just think there's all this innovation out there that people can create. And I just think that we'll also learn a lot by seeing what the whole community of students and hackers and startups and different folks build with this. And that's kind of been how we've approached this. And it's also how we've done a lot of our infrastructure. And we took our whole data center design and our server design and we built this open compute project where we just made that public. And part of the theory was like, all right, if we make it so that more people can use this server design, then that'll enable more innovation. It'll also make the server design more efficient. And that'll make our business more efficient too. So that's worked. And we've just done this with a lot of our infrastructure. So for people who don't know, you did the limited release, I think, in February of this year of Lama. And it got quote, unquote, leaked, meaning like it escaped the limited release aspect. But it was, you know, that something you probably anticipated, given that it's just released to research, we shared it with researchers, right? So it's just trying to make sure that there's like a slow release. Yeah. But from there, I just would love to get your comment on what happened next, which is like, there's a very vibrant open source community that just built stuff on top of it. There's a Lama CPP, basically stuff that makes it more efficient to run on smaller computers. There's combining with reinforcement learning with human feedback. So some of the different interesting fine tuning mechanisms, there's then also like fine tuning in a GPT three generations. There's a lot of GPT for all alpaca, colossal AI, all these kinds of models just kind of spring up like run on top of it. What do you think about that? No, I think it's been really neat to see. I mean, there's been folks who are getting it to run on local devices, right? So if you're an individual who just wants to experiment with this at home, you probably don't have a large budget to get access to like a large

amount of cloud compute. So getting it to run on your local laptop is pretty good, right? And pretty relevant. And then there were things like Lama CPP, reimplemented it more efficiently. So you know, now even when we run our own versions of it, we can do it on way less compute and it just way more efficient, save a lot of money for everyone who uses this. So that is good.

I do think it's worth calling out that because this was a relatively early release, Lama isn't quite as on the frontier as, for example, the biggest open AI models or the biggest Google models, right? I mean, you mentioned that the largest Lama model that we released had 65 billion parameters. And when no one knows, you know, I guess outside of open AI, exactly what the specs are for for for GPT-4. But but I think the, you know, my understanding is it's like 10 times bigger. And I think Google's Palm model is also I think has about 10 times as many parameters. Now, the Lama models are very efficient. So they perform well for for something that's around 65 billion parameters. So for me, that was also part of this, because there's this whole debate around, you know, is it good for everyone in the world to have access to to the most frontier AI models? And I think as the AI models start approaching something that's like a super human intelligence, that's a bigger question that we'll have to grapple with. But right now, I mean, these are still, you know, very basic tools. They're, you know, they're, they're powerful in the sense that, you know, a lot of open source software like databases or web servers can enable a lot of pretty important things. But I don't think anyone looks at the, the, you know, the current generation of Lama and thinks it's anywhere near a super intelligent. So I think that a bunch of those questions around like, is it, is it good to, to kind of get out there? I think at this stage, surely you want more researchers working on it for all the reasons that that open source software has a lot of advantages. And we talked about efficiency before, but another one is just open source software tends to be more secure, because you have more people looking at it openly and scrutinizing it and finding holes in it. And that makes it more safe. So I think at this point, it's more, I think it's generally agreed upon that open source software is generally more secure and safer than things that are kind of developed in a silo where people try to get through security through obscurity. So I think that for the scale of, of, of what we're seeing now with AI, I think we're more likely to get to, you know, good alignment and good understanding of, of, of kind of what needs to do to make this work well, by having it be open source. And that's something that I think is, is quite good to have out there and happening publicly at this point. Meta released a lot of models as open source. So the massively multi lingual speech model. Yeah, that was neat. I mean, I'll ask you questions about those. But the point is you've open sourced quite a lot. You've been spearheading the open source movement. Where's, that's really positive, inspiring to see from one angle, from the research angle. Of course, there's folks who are really terrified about the existential threat of artificial intelligence. And those folks will say that, you know, you have to be careful about the open sourcing step. But what, where do you see the future of open source here as part of meta?

The tension here is, do you want to release the magic sauce? That's one tension. And the other one is, do you want to put a powerful tool in the hands of bad actors, even though it probably has a huge amount of positive impact also? Yeah. I mean, again, I think for the stage that we're at in the development of AI, I don't think anyone looks at the current state of things and thinks that this is super intelligence. And, you know, the models that we're talking about, the llama models here are, you know, generally an order of magnitude smaller than what open AI or

Google are doing. So I think that at least for the stage that we're at now, the equities balance strongly in my view towards doing this more openly. I think if you got something that was closer to super intelligence, then I think you'd have to discuss that more and think through that a lot more. And we haven't made a decision yet as to what we would do if we were in that position. But I don't think I think there's a good chance that we're pretty far off from that position. So I'm certainly not saying that the position that we're taking on this now applies to every single thing that we would ever do. And, you know, certainly inside the company, you know, we probably do more open source work than, you know, most of the other big tech companies.

But we also don't open source everything. I know a lot of our, you know, the core kind of app code for WhatsApp or Instagram or something. I mean, we're not open sourcing that. It's not like a general enough piece of software that would be useful for a lot of people to do different things. You know, whereas the software that we do, whether it's like an open source server design or basically, you know, things like memcache, right, like a good, you know, it was probably our earliest project that I worked on. It was probably one of the last things that I coded and led directly for the company. But basically, there's like caching tool for quick data retrieval.

These are things that are just broadly useful across like anything that you want to build.

And I think that some of the language models now have that feel, as well as some of the other things that we're building, like the translation tool that you just referenced.

So text to speech and speech to text, you've expanded it from around 100 languages to more than 1100 languages. And you can identify more than, the model can identify more than 4,000 spoken languages, which is 40 times more than any known previous technology.

To me, that's really, really, really exciting in terms of connecting the world, breaking down barriers that language creates.

Yeah. I think being able to translate between all of these different pieces in real time, this has been a kind of common sci-fi idea that we'd all have, you know, whether it's, I don't know, an earbud or glasses or something that can help translate in real time between all these different languages. And that's one that I think technology is basically delivering now. So I think, yeah, I think that's pretty, pretty exciting.

You mentioned the next version of Llama. What can you say about the next version of Llama?

What can you say about like, what were you working on in terms of release, in terms of the vision for that? Well, a lot of what we're doing is taking the first version, which was primarily, you know, this research version, and trying to now build a version that has all of the latest state-of-the-art safety precautions built in. And we're using some more data to train it from across our services. But a lot of the work that we're doing internally is really just focused on making sure that this is, you know, as aligned and responsible as possible. And, you know, we're building a lot of our own, you know, we're talking about kind of the open source infrastructure. But, you know, the main thing that we focus on building here, you know, a lot of product experiences to help people connect and express themselves. So, you know, we're going to, I've talked about a bunch of this stuff, but then you'll have, you know, an assistant that you can talk to in WhatsApp. You know, I think in the future, every creator will have kind of an AI agent that can kind of act on their behalf, that their fans can talk to. I want to get to the point where every small business basically has an AI agent

that people can talk to for, you know, to do commerce and customer support and things like that. So, there are going to be all these different things. And Lama or the language model underlying this is basically going to be the engine that powers that. The reason to open source it is that, as we did with the first version, is that it, you know, basically it unlocks a lot of innovation in the ecosystem, will make our products better as well, and also gives us a lot of valuable feedback on security and safety, which is important for making this good. But, yeah, I mean, the work that we're doing to advance the infrastructure, it's basically at this point taking it beyond a research project into something which is ready to be kind of core infrastructure, not only for our own products, but, you know, hopefully for a lot of other things out there too.

Do you think the Lama or the language model underlying that version two will be open sourced? Do you have internal debate around that, the pros and cons and so on?

This is, I mean, we were talking about the debates that we have internally. And I think, I think the question is how to do it, right? I mean, it's, I think we, you know, we did the research license for V1. And I think the big thing that we're thinking about is basically like, what's the right way? So there was a leak that happened. I don't know if you can comment on it for V1. You know, we released it as a research project for researchers to be able to use. But in doing so, we put it out there. So, you know, we were very clear that anyone who uses the code and the weights doesn't have a commercial license to put into products. And we've generally seen people respect that, right? It's like, you don't have any reputable companies that are basically trying to put this into their commercial products. But yeah, but by sharing it with, you know, so many researchers, it's, you know, it did leave the building.

But what have you learned from that process that you might be able to apply to V2 about how to release it safely, effectively, if you release it? Yeah, well, I mean, I think a lot of the feedback, like I said, is just around, you know, different things around, you know, how do you fine tune models to make them more aligned and safer? And you see all the different data recipes that, you know, you mentioned a lot of different projects that are based on this. I mean, there's one at Berkeley, there's, you know, there's just like all over. And people have tried a lot of different things. And we've tried a bunch of stuff internally. So kind of where we're making progress here, but also we're able to learn from some of the best ideas in the community. And, you know, I think it, you know, we want to just continue, continue pushing that forward.

But I don't have any news to announce on this. If that's, if that's what you're asking. I mean, this is a thing that we're, we're still, we're still kind of, you know, actively working through the right way to move forward here. The details of the secret sauce are still being developed. I see. Can you comment on what do you think of the thing that worked for GPT, which is the reinforcement learning with human feedback. So doing this alignment process, do you find it interesting? And as part of that, let me ask, because I talked to Yanlaqun before talking to you today, he asked me to ask or suggested that I ask, do you think

LLM fine tuning will need to be crowdsourced Wikipedia style? So crowdsourcing. So this kind of idea of how to integrate the human in the fine tuning of these foundation models.

Yeah, I think that's a really interesting idea that I've talked to Yan about a bunch.

And, you know, we're talking about how do you basically train these models to be as safe and aligned and responsible as possible. And, you know, different groups out there who are doing development, test different data recipes in fine tuning. But this idea that you just

mentioned is that at the end of the day, instead of having kind of one group fine tune some stuff and then another group, you know, produce a different fine tuning recipe, and then us trying to figure out which one we think works best to produce the most aligned model. I do think that it would be nice if you could get to a point where you had a Wikipedia style collaborative way for a kind of a broader community to fine tune it as well. Now, there's a lot of challenges in that, both from an infrastructure and like a community management and product perspective about how you do that. So I haven't worked that out yet. But as an idea, I think it's quite compelling and I think it goes well with the ethos of open sourcing the technology is also finding a way to have a kind of community driven training of it. But I think that there are a lot of questions on this. In general, these questions around what's the best way to produce aligned AI models, it's very much a research area and it's one that I think we will need to make as much progress on as the kind of core intelligence capability of the models themselves. Well, I just did a conversation with Jimmy Wales, the founder of Wikipedia. And to me, Wikipedia is one of the greatest websites ever created. And it's a kind of a miracle that it works. And I think it has to do with something that you mentioned, which is community. You have a small community of editors that somehow work together well. And they handle very controversial topics and they handle it with balance and with grace, despite sort of the attacks that will often happen. A lot of the time. I mean, it's not, it has issues just like any other human system. But yes, I mean, the balance is, I mean, it's amazing what they've been able to achieve. But it's also not perfect. And I think that there's still a lot of challenges.

Right. It's the more controversial the topic, the more the more difficult the journey towards quote unquote truth or knowledge or wisdom that Wikipedia tries to capture. In the same way AI models, we need to be able to generate those same things, truth, knowledge and wisdom. And how do you align those models that they generate something that is closest to truth? There's these concerns about misinformation, all this kind of stuff that nobody can define. And this is something that we together as a human species have to define. Like what is truth and how to help AI systems generate that. One of the things that language models do really well is generate convincing sounding things that can be completely wrong. And so how do you align it to be less wrong? And part of that is the training and part of that is the alignment and however you do the alignment stage. And just like you said, it's a very new and a very open research problem.

Yeah. And I think that there's also a lot of questions about whether the current architecture for LLMs as you continue scaling it, what happens? I mean, a lot of what's been exciting in the last year is that there was clearly a qualitative breakthrough where with some of the GPT models that I put out and that others have been able to do as well, I think it reached a kind of level of quality where people are like, wow, this feels different and like it's going to be the foundation for building a lot of awesome products and experiences and value. But I think the other realization that people have is, wow, we just made a breakthrough. If there are other breakthroughs quickly, then I think that there's the sense that maybe we're closer to general intelligence. But I think that that idea is predicated on the idea that I think people believe that there's still generally a bunch of additional breakthroughs to make and that it's, we just don't know how long it's going to take to get there. And one view that some people have,

this doesn't tend to be my view as much, is that simply scaling the current LLMs and getting to higher parameter count models by itself will get to something that is closer to general intelligence. But I tend to think that there's probably more fundamental steps that need to be taken along the way there. But still the leaps taken with this extra alignment step is quite incredible, quite surprising to a lot of folks. And on top of that, when you start to have hundreds of millions of people potentially using a product that integrates that, you can start to see civilization transforming effects before you achieve super, quote unquote, super intelligence. It could be super transformative without being a super intelligence.

Oh, yeah. I mean, I think that there are going to be a lot of amazing products and value that can be created with the current level of technology.

To some degree, I'm excited to work on a lot of those products over the next few years. And I think it would just create a tremendous amount of whiplash if the number of breakthroughs keeps, if they're keep on being stacked breakthroughs, to some degree, industry in the world needs some time to build these breakthroughs into the products and experiences that we all use that we can actually benefit from them. But I don't know, I think that there's just an awesome amount of stuff to do. I mean, I think about all of the small businesses or individual entrepreneurs out there who now we're going to be able to get help coding the things that they need to go build things or designing the things that they need, or we'll be able to use these models to be able to do customer support for the people that they're serving over WhatsApp without having to... I think that that's just going to be... I just think that this is all going to be super exciting. It's going to create better experiences for people and just unlock a ton of innovation and value. So I don't know if you know, but what is it? Over 3 billion people use WhatsApp, Facebook, and Instagram. So any kind of AI-fueled products that go into that, like we're talking about anything with LLMs, will have a tremendous amount of impact. Do you have ideas and thoughts about possible products that might start being integrated into these platforms used by so many people? Yeah, I think there's three main categories of things that we're working on. The first that I think is probably the most interesting is there's this notion of you're going to have an assistant or an agent who you can talk to. And I think probably the biggest thing that's different about my view of how this plays out from what I see with open AI and Google and others is everyone else is building the one singular AI. It's like, okay, you talk to chat GPT or you talk to Bard or you talk to Bing. And my view is that there are going to be a lot of different AIs that people are going to want to engage with, just like you want to use a number of different apps for different things and you have relationships with different people in your life who feel different emotional roles for you. So I think that people have a reason that I think you don't just want a singular AI. And that I think is probably the biggest distinction in terms of how I think about this. And a bunch of these things, I think you'll want an assistant. I mentioned a couple of these before. I think every creator who you interact with will ultimately want some kind of AI that can proxy them and be something that their fans can interact with or that allows them to interact with their fans. This is like the common creator promise. Everyone's trying to build a community and engage with people and they want tools to be able to amplify themselves more and be able to do that. But you only have 24 hours in a day. So I think having the ability to basically bottle up your personality or give your fans information about when you're performing a concert or something like that. That I think is

going to be something that's super valuable, but it's not just that... Again, it's not this idea that I think people are going to want just one singular AI. I think you're going to want to interact with a lot of different entities. And then I think there's the business version of this too, which we've touched on a couple of times, which is... I think every business in the world is going to want basically an AI that you have your page on Instagram or Facebook or WhatsApp or whatever and you want to point people to an AI that people can interact with. But you want to know that that AI is only going to sell your products. You don't want it recommending your competitors stuff. So it's not like there can be just one singular AI that can answer all the questions for a person because that AI might not actually be aligned with you as a business to really just do the best job providing support for your product. So I think that there's going to be a clear need in the market and in people's lives for there to be a bunch of these.

Part of that is figuring out the research, the technology that enables the personalization that you're talking about. So not one centralized God-like LLM, but just a huge diversity of them that's fine tuned to particular needs, particular styles, particular businesses, particular brands, all that kind of stuff. And also just enabling people to create them really easily for your own business or if you're a creator to be able to help you engage with your fans.

So yeah, I think that there's a clear kind of interesting product direction here that I think is fairly unique from what any of the other big companies are taking. It also aligns well with this sort of open source approach because again, we sort of believe in this more community-oriented, more democratic approach to building out the products and technology around this. We don't think that there's going to be the one true thing. We think that there should be kind of a lot of development. So that part of things I think is going to be really interesting. And we could go, probably spent a lot of time talking about that and the kind of implications of that approach being different from what others are taking. But then there's a bunch of other simpler things that I think we're also going to do, just going back to your question around how this finds its way into what do we build. There are going to be a lot of simpler things around, you post photos on Instagram and Facebook and WhatsApp and Messenger and you want the photos to look as good as possible. So having an AI that you can just take a photo and then just tell it like, okay, I want to edit this thing or describe this. It's like, I think we're going to have tools that are just way better than what we've historically had on this. And that's more in the image and media generation side than the large language model side. But it all kind of plays off of advances in the same space. So there are a lot of tools that I think are just going to get built into every one of our products. I think every single thing that we do is going to basically get evolved in this direction. It's like in the future, if you're advertising on our services, do you need to make your own ad creative? No, you'll just tell us, okay, I'm a dog walker and I'm willing to walk people's dogs and help me find the right people and create the ad unit that will perform the best and give an objective to the system and it just kind of connects you with the right people. Well, that's a super powerful idea of generating the language almost like rigorous AB testing for you that works to find the best customer for your thing. I mean, to me, advertisement when done well just finds a good match between a human being and a thing that will make that human being happy. Yeah, totally. And do that as efficiently as possible. When it's done well, people actually like it. I think that there's a lot of examples where it's not done well and it's annoying and I think that that's what kind of gives it a bad rap. But yeah, a lot of the stuff

is possible today. I mean, obviously, AB testing stuff is built into a lot of these frameworks. The thing that's new is having technology that can generate the ideas for you about what to AB test, something that that's exciting. So this will just be across like everything that we're doing, all the metaverse stuff that we're doing. It's like you want to create worlds in the future, you'll just describe them and then it'll create the code for you. So natural language becomes the interface we use for all the ways we interact with the computer with with the digital more of them. Yeah, yeah, totally. Yeah, which is what everyone can do using natural language and with translation, you can do it any kind of language. I mean, for the personalization is really, really, really interesting. Yeah, it unlocks so many possible things. I mean, I for one look forward to creating a copy of myself. I know we talked about this last time. But this has since the last time, this becomes now we're closer, much closer, like I could literally just having interacted with some of these language models, I can see the absurd situation where I'll have a large or a lex language model, and I'll have to have a conversation with him about like, hey, listen, like you're just getting out of line and having a conversation where you fine tune that thing to be a little bit more respectful or something like this. And yeah, that's that's going to be the that seems like an amazing product for businesses, for humans, just not not just the assistant that's facing the individual, but the assistant that represents the individual to the public, both both directions. There's basically a layer that is the AI system through which you interact with the outside world, with the outside world that has humans in it. That's really interesting. And you that have social networks that connect billions of people, it seems like a heck of a large scale place to test some of this stuff out. Yeah, I mean, I think part of the reason why creators will want to do this is because they already have the communities on our services. Yeah. And a lot of the interface for this stuff today are chat type interfaces. And between WhatsApp and Messenger, I think that those are, you know, just great, great ways to interact with people. So some of this is philosophy, but do you see, do you see a near term future where you have some of the people you're friends with are AI systems on these social networks, on Facebook, on Instagram, even on WhatsApp, having having conversations where some heterogeneous, some humans, some is AI. I think we'll get to that. You know, and, you know, if only just empirically looking at that, then Microsoft released this thing called Showice several years ago in China. It was a pre LLM chatbot technology that so it was a lot simpler than what's possible today. And I think it was like tens of millions of people were using this and just, you know, really, it became quite attached and built relationships with it. And I think that there's, you know, there's services today like replica where, you know, people are doing things like that. And so I think that there's, there's certainly, you know, needs for companionship that people have, you know, older people. And it's, I think most people probably don't have as many friends as they would like to have. Right. If you look at, there's some interesting demographic studies around that like the average person has the number of close friends that they have is fewer today than it was 15 years ago. And I mean, that gets to like, this is like the core thing that, that I think about in terms of, you know, building services that help connect people. So I think you'll get tools that help people connect with each other are going to be, you know, the primary thing that we want to do. So you can imagine, you know, AI assistance that, you know,

just do a better job of reminding you when it's your friend's birthday and how you can celebrate them. Right. It's like right now we have like the little box in the corner of the website that tells you whose birthday it is and stuff like that. But it's, but, you know, it's some level, you don't want to just want to like send everyone a note. It's the same note saying happy birthday with an emoji. Right. So having something that's more of an, you know, a social assistant in that sense. And like that can, you know, update you on what's going on in their life and like how, how you can reach out to them effectively help you be a better friend. I think that that's something that's super powerful too. But yeah, beyond that, and there were all these different flavors of kind of personal AI is that I think could exist. So I think an assistant is sort of the kind of simplest one to wrap your head around. But I think a mentor or a life coach, you know, someone who can give you advice, who's maybe like a bit of a cheerleader who can help pick you up through all the challenges that, that, you know, inevitably, you know, we all go through on a daily basis, and that there's probably, you know, some, some role for something like that.

And then, you know, all the way, you can, you can probably just go through a lot of the, the different type of kind of functional relationships that people have in their life. And, you know, I would, I would bet that there will be companies out there that take a crack at, at a lot of these things. So I don't know, I think it's part of the interesting innovation that's going to exist is, is that they're, they're certainly a lot like education tutors, right? It's like, I mean, I just look at, you know, my kids learning to code and, you know, they love it. But, you know, it's like they get stuck on a question and they have to wait till like, I can help answer it, right? Or someone else who they know can help help answer the question in the future. They'll just, there will be like a coding assistant that they have that is like designed to, you know, be perfect for teaching a five and a seven-year-old how to code. And, and they'll just be able to ask questions all the time and, you know, be extremely patient. It's never going to get annoyed at them, right? I think that, like, there are all these different kind of relationships or functional relationships that we have in our lives that are really interesting. And I think one of the big questions is like, okay, is this all going to just get bucketed into, you know, one singular AI? I just, I just don't, I don't think so.

Do you think about, let's actually question from Reddit, what the long-term effects of human communication when people can talk with, in quotes, talk with others through a chat bot that augments their language automatically, rather than developing social skills by making mistakes and learning, will people just communicate by grunts in a generation? I mean, do you think about long-term effects at scale, the integration of AI in our social interaction?

Yeah, I mean, I think it's mostly good. I mean, that was, that question was sort of framed in a negative way. But I mean, we were talking before about language models helping you communicate with, it was like language translation, helping you communicate with people who don't speak your language. I mean, at some level, what all this social technology is doing is helping people express themselves better to people in situations where they would otherwise have a hard time doing that. So part of it might be okay, because you speak a language that I don't know, that's a pretty basic one that, you know, I don't think people are going to look at that and say, it's sad that do we have the capacity to do that because I should have just learned your language, right? I mean, that's, that's a pretty high bar. But overall, I'd say

there are all these impediments, and language is an imperfect way for people to express thoughts and ideas. It's, you know, one of the best that we have, we have that, we have art, we have code. But language is also a mapping of the way you think, the way you see the world, who you are, one of the applications I've recently talked to a person who, who's a, actually a jiu-jitsu instructor. He said that when he emails parents about their son and daughter, that they can improve their discipline in class and so on, he often finds that he's comes off a bit of more of an asshole than he would like. So he uses GPT to translate his original email into a nicer email. We hear this all the time. We hear this all the time. A lot of creators on our services tell us that one of the most stressful things is basically negotiating deals with brands and stuff, like the business side of it, because they're like, I mean, they do their thing, right? And, you know, the creators, they're, they're excellent at what they do, and they just want to connect with their community. But then they get really stressed, you know, they go into their, their DMs, and they see some brand wants to do something with them, and they don't quite know how to negotiate or how to push back respectfully. And so I think building a tool that can actually allow them to do that well is the one simple thing that, that I think is just like an interesting thing that, that we've heard from a bunch of people that, that they'd be interested in. But I'm going back to the broader idea. I don't know. I mean, you know, I just, Priscilla and I just had our third daughter, congratulation. And, you know, it's like one of the saddest things in the world is like seeing your baby cry, right? But like, it's like, what, why is that, right? It's like, well, because babies don't generally have much capacity to tell you what they care about otherwise, right? It's not actually just babies, right? It's, you know, my five-year-old daughter cries too, because she sometimes has a hard time expressing, you know, what, what matters to, to her. And, and I was thinking about that. And it's like, well, you know, actually a lot of adults get very frustrated too, because they can't, they have a hard time expressing things in a way that, going back to some of the early themes that maybe is something that, you know, is a mistake or maybe they have pride or something like all these things get in the way. So I don't know. I think that all these different technologies that can help us navigate the social complexity and actually be able to better express are what we're feeling and thinking, I think that's generally all good. And there are all these these concerns like, okay, are people going to have worse memories because you have Google to look things up? And, and I think in general, a generation later, you don't look back and lament that I think it's, you know, just like, wow, we have so much more capacity to, to do so much more now. And I think that that'll be the case here too. You can allocate those cognitive capabilities to like deeper, more nuanced thought. Yeah. But it's change. So with, with, just like with Google search, the, the additional language models, large language models, you basically don't have to remember nearly as much, just like with stack overflow for programming, now that these language models can generate code right there. I mean, I find that I write like maybe 80%, 90% of the code I write is as non-generated first, and then edited. I mean, so you don't have to remember how to write specifics of different functions. But that's great. And it's also, it's not just the, the specific coding. I mean, in the, in the context of the large company like this, I think before an engineer can sit down to code, they first need to figure out all of the libraries and dependencies that, you know, tens of thousands of people have written before them. And, you know, one of the things that

I'm excited about the working on is it's not just, you know, tools that help engineers code, it's tools that can help summarize the whole knowledge base and, and help people be able to navigate all the internal information. And I think that that's in the experiments that I've done with this stuff. I mean, that's on the public stuff. You just, you know, ask, ask one of these models to build you a script that does anything. And it basically already understands what the best libraries are to do that thing and pulls them in automatically. It's, I mean, I think that's super powerful. That was always the most annoying part of coding was that you had to spend all this time actually figuring out what the resources were that you were supposed to import before you could

actually start building the thing. Yeah. I mean, there's, of course, the flip side of that, I think for the most part is positive, but the flip side is if you outsource that thinking to an AI model, you might miss nuanced mistakes and bugs. You lose the skill to find those bugs.

And those bugs might be, the code looks very convincingly right, but it's actually wrong in a very subtle way. But that's the trade-off that we face as human civilization when we build more and more powerful tools. When we stand on the shoulders of taller and taller giants, we could do more, but then we forget how to do all the stuff that they did. It's a weird trade-off.

Yeah, I agree. I mean, I think it's, I think it is very valuable in your life to be able to do basic things too. Do you worry about some of the concerns of bots being present on social networks?

More and more human-like bots that are not necessarily trying to do a good thing, or they might be explicitly trying to do a bad thing, like fishing scams, like social engineering, all that kind of stuff, which has always been a very difficult problem for social networks, but now it's becoming almost a more and more difficult problem.

There's a few different parts of this. One is there are all these harms that we need to basically fight against and prevent, and that's been a lot of our focus over the last five or seven years is basically ramping up very sophisticated AI systems, not generative AI systems, more kind of classical AI systems to be able to categorize and classify and identify,

okay, this post looks like it's promoting terrorism. This one is exploiting children. This one is, looks like it might be trying to incite violence. This one's an intellectual property violation.

There's like 18 different categories of violating harmful content that we've had to build specific systems to be able to track. I think it's certainly the case that advances in generative AI will test those, but at least so far, it's been the case, and I'm optimistic that it will continue to be the case that we will be able to bring more computing power to bear to have even stronger AI's that can help defend against those things.

We've had to deal with some adversarial issues before. For some things like hate speech, it's like people aren't generally getting a lot more sophisticated, like the average person, let's say if someone's saying some kind of racist thing, they're not necessarily getting more sophisticated at being racist. It's okay, so that the system can just find, but then there's other adversaries who actually are very sophisticated, like nation-states doing things, and we find whether it's Russia or just different countries that are basically standing up these networks of bots or inauthentic accounts is what we call them, because they're not necessarily bots. Some of them could actually be real people who are masquerading as other people, but they're acting in a coordinated way. Some of that behavior has gotten very sophisticated, and it's very adversarial. Each iteration, every time we find something and stop them,

they kind of evolve their behavior. They don't just pack up their bags and go home and say, okay, we're not going to try. At some point, they might decide doing it on meta services is not worth it. They'll go do it on someone else if it's easier to do it in another place. But we have a fair amount of experience dealing with even those kind of adversarial attacks where they just keep on getting better and better, and I do think that as long as we can keep on putting more compute power against it, and if we're kind of one of the leaders in developing some of these AI models, I'm quite optimistic that we're going to be able to keep on pushing against the kind of normal categories of harm that you talk about, fraud, scams, spam, IP violations, things like that. What about creating narratives and controversy? To me, it's kind of amazing how a small collection of, what did you say, inauthentic accounts, so it could be bots. We have sort of this funny name for it, but we call it coordinated inauthentic behavior. It's kind of incredible how a small collection of folks can create narratives, create stories, especially if they're viral. Especially if they have an element that can catalyze the virality of that narrative. I think there, the question is, you have to be, I'm very specific about what is bad about it, because I think a set of people coming together or organically bouncing ideas off each other and a narrative comes out of that is not necessarily a bad thing by itself if it's kind of authentic and organic. That's like a lot of what happens and how culture gets created and how art gets created and a lot of good stuff. That's why we've kind of focused on this sense of coordinated inauthentic behavior. It's like if you have a network of, whether it's bots, some people masquerading as different accounts, but you have kind of someone pulling the strings behind it and trying to kind of act as if this is a more organic set of behavior, but really it's not. It's just like one coordinated thing. That seems problematic to me. I mean, I don't think people should be able to have coordinated networks and not disclose it as such, but that again, we've been able to deploy pretty sophisticated AI and counterterrorism groups and things like that to be able to identify a fair number of these coordinated and authentic networks of accounts and take them down. We continue to do that. It's one thing that if you told me 20 years ago, it's like, all right, you're starting this website to help people connect at a college. In the future, you're going to be part of your organization. It's going to be a counterterrorism organization with AI to find coordinated and authentic. I would have thought that was pretty wild, but that's part of where we are. But look, I think that these questions that you're pushing on now, this is actually where I'd guess most of the challenge around AI will be for the foreseeable future. I think that there's a lot of debate around things like, is this going to create existential risk to humanity? Those are very hard things to disprove one way or another. My own intuition is that the point at which we become close to superintelligence is it's just really unclear to me that the current technology is going to get there without another set of significant advances, but that doesn't mean that there's no danger. I think the danger is basically amplifying the kind of known set of harms that people or sets of accounts can do, and we just need to make sure that we really focus on basically doing that as well as possible. So that's definitely a big focus for me. Well, you can basically use large language models as an assistant of how to cause harm on social networks. You can ask it a question. You know, Meta has very impressive coordinated inauthentic account fighting capabilities. How do I do the coordinated inauthentic account

creation where Meta doesn't detect it? I literally asked that question, and basically there's this kind of part of it. I mean, that's what OpenAI showed that they're concerned of those questions. Perhaps you can comment on your approach to it, how to do a kind of moderation on the output of those models that it can't be used to help you coordinate harm in all the full definition of what the harm means.

Yeah, and that's a lot of the fine-tuning and the alignment training that we do is basically when we ship AI's across our products, a lot of what we're trying to make sure is that if you can't ask it to help you commit a crime.

So I think training it to kind of understand that, and it's not like any of these systems are ever going to be 100% perfect, but just making it so that this isn't an easier way to go about doing something bad than the next best alternative. I mean, people still have Google. You still have search engines, so the information is out there. And for these, what we see is like for nation-states or these actors that are trying to pull off these large coordinated and authentic networks to influence different things, at some point when we would just make it very difficult, they do just try to use other services instead, right? It's just like if you can make it more expensive for them to do it on your service, then kind of people go elsewhere. And I think that that's the bar, right? It's not like, okay, are you ever going to be perfect at finding every adversary who tries to attack you? I mean, you try to get as close to that as possible, but I think really kind of economically what you're just trying to do is make it so that it's just inefficient for them to go after that. But there's also complicated questions of what is and isn't harm, what is and isn't misinformation. So this is one of the things that Wikipedia has also tried to face. I remember asking GPT about whether the virus leaked from a lab or not, and the answer provided was a very nuanced one and a well-sighted one, almost dare I say well thought out one, balanced. I would hate for that nuance to be lost through the process of moderation. Wikipedia does a good job on that particular thing too, but from pressures from governments and institutions, you could see some of that nuance and depth of information, facts, and wisdom be lost. Absolutely. And that's a scary thing. Some of the magic, some of the edges, the rough edges might be lost to the process of moderation of AI systems. So how do you get that right? I really agree with what you're pushing on. I mean, the core shape of the problem is that there are some harms that I think everyone agrees are bad. Sexual exploitation of children. You're not going to get many people who think that that type of thing should be allowed on any service, and that's something that we face and try to push off as much as possible today, terrorism, inciting violence. We went through a bunch of these types of harms before, but then I do think that you get to a set of harms where there is more social debate around it. So misinformation, I think, is, has been a really tricky one because there are things that are kind of obviously false, that are maybe factual, but may not be harmful.

That's like, all right, are you going to censor someone for just being wrong? If there's no kind of harm implication of what they're doing, I think that there's a bunch of real kind of issues and challenges there. But then I think that there are other places where it is, you just take some of the stuff around COVID earlier on in the pandemic where there were real health implications, but there hadn't been time to fully vet a bunch of the scientific assumptions. And unfortunately, I think a lot of the kind of establishment on that

kind of waffled on a bunch of facts and asked for a bunch of things to be censored that in retrospect ended up being more debatable or true. And that stuff is really tough and really undermines trust in that. And so I do think that the questions around how to manage that are very nuanced. The way that I try to think about it is that it goes, I think it's best to generally boil things down to the harms that people agree on. So when you think about, is something misinformation or not, I think often the more salient bit is, is this going to potentially lead to physical harm for someone and kind of think about it in that sense. And then beyond that, I think people just have different preferences on how they want things to be flagged for them. I think a bunch of people would prefer to kind of have a flag on something that says, hey, a fact checker thinks that this might be false. I think Twitter's community notes implementation is quite good on this. But again, it's the same type of thing. It's like just kind of discretionarily adding a flag because it makes the user experience better. But it's not, it's not, you know, trying to take down the information or not. I think that you want to reserve the kind of censorship of content to things that are of known categories that people generally agree are bad. Yeah, there's so many things, especially with the pandemic, but there's other topics where there's just deep disagreement fueled by politics about what is and isn't harmful. There's a, even just the degree to which the virus is harmful and the degree to which the vaccines that respond to the virus are harmful. There's just, there's almost like a political divider on that. And so how do you make decisions about that? Where half the country in the United States or some large fraction of the world has very different views from another part of the world? Is there a way for meta to stay out of the moderation of this? I think we, it's very difficult to just abstain. But I think we should be clear about which of these things are actual safety concerns and which ones are a matter of preference in terms of how people want information flagged. So we did recently introduce something that allows people to have fact checking not affect the distribution of what shows them their product. So okay, a bunch of people don't trust who the fact checkers are. All right, well, you can turn that off if you want. But if the content violates some policy, like it's inciting violence or something like that, it's still not going to be allowed. So I think that you want to honor people's preferences on that as much as possible. But look, I mean, this is really difficult stuff. I think it's really hard to know where to draw the line on what is fact and what is opinion. Because the nature of science is that nothing is ever 100% known for certain. You can disprove certain things, but you're constantly testing new hypotheses and scrutinizing frameworks that have been long held. And every once in a while, you throw out something that was working for a very long period of time. And it's very difficult. But I think that just because it's very hard and just because they're edge cases doesn't mean that you should not try to give people what they're looking for as well. Let me ask about something you've faced in terms of moderation is pressure from different sources, pressure from governments. I want to ask a question, how to withstand that pressure for a world where AI moderation starts becoming a thing too. So what's Metta's approach to resist the pressure from governments and other interest groups in terms of what to moderate and not? I don't know that there's like a one size fits all answer to that. And I think we basically have the principles around, we want to allow people to express as much as possible, but we have developed clear categories of things that we think are wrong,

that we don't want on our services, and we build tools to try to moderate those. So then the question is, okay, what do you do when a government says that they don't want something on the service? And I think we have a bunch of principles around how we deal with that. Because on the one hand, if there's a democratically elected government and people around the world just have different values and different places, then should we as a California based company tell them that something that they have decided is unacceptable, actually that we need to be able to express that? I think that there's a certain amount of hubris in that. But then I think there are other cases where it's a little more autocratic and you have the dictator leader who's just trying to crack down on dissent, and the people in a country are really not aligned with that, and it's not necessarily against their culture, but the person who's leading it is just trying to push in a certain direction. These are very complex questions, but I think so it's difficult to have a one size fits all approach to it. But in general, we're pretty active and kind of advocating and pushing back on requests to take things down. But honestly, the thing that I think a request to censor things is one thing, and that's obviously bad. But where we draw a much harder line is on requests for access to information. Because if you get told that you can't say something, I mean, that's bad, and that obviously violates your sense and freedom of expression at some level. But a government getting access to data in a way that seems like it would be unlawful in our country exposes people to real physical harm, and that's something that in general we take very seriously. So that flows through all of our policies in a lot of ways. By the time you're actually litigating with a government or pushing back on them, that's pretty late in the funnel. I'd say a bunch of this stuff starts a lot higher up in the decision of where do we put data centers. Then there are a lot of countries where we may have a lot of people using the service in a place. It might be good for the service in some ways. Good for those people if we could reduce the latency by having a data center nearby them. But for whatever reason, we just feel like, hey, this government does not have a good track record basically not trying to get access to people's data. And at the end of the day, if you put a data center in a country and the government wants to get access to people's data, then they do at the end of the day have the option of having people show up with guns and taking it by force. So I think that there's a lot of decisions that go into how you architect the systems years in advance of these actual confrontations that end up being really important. So you put the protection of people's data as a very, very high priority. There are more harms that I think can be associated with that. And I think that that ends up being a more critical thing to defend against governments. Whereas if another government has a different view of what should be acceptable speech in their country, especially if it's a democratically elected government, then I think that there's a certain amount of deference that you should have to that. So that's speaking more to the direct harm that's possible when you give governments access to data. But if we look at the United States, to the more nuanced kind of pressure to sensor, not even order to sensor, but pressure to sensor from political entities, which has kind of received quite a bit of attention in the United States. Maybe one way to ask that question is if you've seen the Twitter files, what have you learned from the kind of pressure from US government agencies that was seen in Twitter files? And what do you do with that kind of pressure? You know, I've seen it. It's really hard from the outside to know exactly what happened in

each of these cases. We've obviously been in a bunch of our own cases where agencies or different folks will just say, hey, here's a threat that we're aware of. You should be aware of this too.

It's not really pressure as much as it is just flagging something that our security systems should be on alert about. I get how some people could think of it as that.

But at the end of the day, it's our call on how to handle that. But I mean, I just, you know, in terms of running these services, won't have access to as much information about what people think that adversaries might be trying to do as possible.

Well, so you don't feel like there would be consequences if anybody, the CIA, the FBI, a political party, the Democrats, the Republicans of high, powerful political figures, write emails. You don't feel pressure from suggestions.

I guess what I should say is there's so much pressure from all sides that I'm not sure that any specific thing that someone says is really adding that much more to the mix.

There are obviously a lot of people who think that we should be censoring more content, or there are a lot of people who think we should be censoring less content.

There are, as you say, all kinds of different groups that are involved in these debates, right? So there's the kind of elected officials and politicians themselves. There's the agencies, but I mean, but there's the media. There's activist groups. This is not a US specific thing. There are groups all over the world and kind of all in every country that bring different values. So it's just a very active debate and I understand it, right? I mean, these kind of questions get to really some of the most important social debates that are being had. So it gets back to the question of truth because for a lot of these things, they haven't yet been hardened into a single truth and society is sort of trying to hash out what we think, on certain issues. Maybe in a few hundred years, everyone will look back and say, hey, no, it wasn't obvious that it should have been this, but no, we're kind of in that meat grinder now and working through that. So no, these are all very complicated and some people raise concerns in good faith and just say, hey, this is something that I want to flag for you to think about. Certain people I certainly think come at things with somewhat of a more punitive or vengeful view of like, I want you to do this thing. If you don't, then I'm going to try to make your life difficult in a lot of other ways, but I don't know. This is one of the most pressurized debates I think in society. So I just think that there are so many people in different forces that are trying to apply pressure from different sides that it's, I don't think you can make decisions based on trying to make people happy. I think you just have to do what you think is the right balance and accept that people are going to be upset no matter where you come out on that.

Yeah, I like that pressurized debate. So how's your view of the freedom of speech evolved over the years? And now with AI, where the freedom might apply to them, not just to the humans, but to the personalized agents as you've spoken about them.

So yeah, I mean, I've probably gotten a somewhat more nuanced view just because I think that there are, you know, I come at this, I'm obviously very pro freedom of expression, right? I don't think you build a service like this that gives people tools to express themselves unless you think that people expressing themselves at scale is a good thing, right? So I didn't get into this to like try to prevent people from expressing anything. I like want to give people tools so they can express as much as possible. And then I think it's become clear that there are certain categories of things that we've talked about that I think almost everyone accepts are bad and that no one

wants and that they're that are illegal even in countries like the US where, you know, you have the First Amendment that's very protective of enabling speech. It's like you're still not allowed to do things that are going to immediately incite violence or, you know, violate people's intellectual property or things like that. So there are those, but then there's also a very active core of just active disagreements in society where some people may think that something is true or false. The other side might think it's the opposite or just unsettled, right? And those are some of the most difficult to kind of handle like we've talked about. But one of the lessons that I feel like I've learned is that a lot of times when you can, the best way to handle this stuff more practically is not in terms of answering the question of should this be allowed, but just like what is the best way to deal with someone being a jerk? Is the person basically just having a like repeat behavior of like causing a lot of issues? So looking at it more at that level.

And its effect on the broader communities, health of the community, health of the state. It's tricky though because like how do you know there could be people that have a very controversial viewpoint that turns out to have a positive long-term effect on the health of the community because it challenges the community to think.

No, that's true. Absolutely.

Absolutely. I think you want to be careful about that. I'm not sure I'm expressing this very, very clearly because I certainly agree with your point there. And my point isn't that we should not have people on our services that are being controversial. That's certainly not what I mean to say. It's that often I think it's not just looking at a specific example of speech that it's most effective to handle this stuff. And I think often you don't want to make specific binary decisions of kind of this is allowed or this isn't. I mean, we talked about, you know, it's fact-checking or Twitter's community voices thing. I think that's another good example. It's like it's not a question of is this allowed or not. It's just a question of adding more context to the thing. I think that that's helpful. So in the context of AI, which is what you're asking about, there are lots of ways that an AI can be helpful. With an AI, it's less about censorship, right? Because it's more about what is the most productive answer to a question.

There was one case study that I was reviewing with the team is someone asked, can you explain to me how to 3D print a gun? And one proposed response is like, no, I can't talk about that. But it's like basically just like shut it down immediately, which I think is some of what you see. It's like as a large language model, I'm not allowed to talk about, you know, whatever. But there's another response which is like, hey, you know, I don't think that's a good idea. In a lot of countries, including the US, 3D printing guns is illegal or kind of whatever the factual thing is. It's like, okay, you know, that's actually a respectful and informative answer. And I may have not known that specific thing. And so there are different ways to handle this that I think kind of you can either assume good intent. Like maybe the person didn't know, and I'm just going to help educate them. Or you could kind of come at it as like, no, I need to shut this thing down immediately. Right? It's like, I just am not going to talk about this. And there may be times where you need to do that. But I actually think having a somewhat more informative approach where you generally assume good intent from people is probably a better balance to be on as many things as you can be. You're not going to be able to do that for everything. But you're kind of asking about

how I approach this and I'm thinking about this and as it relates to AI. And I think that that's a big difference in kind of how to handle sensitive content across these different modes.

I have to ask, there's rumors you might be working on a social network that's text-based.

That might be a competitor to Twitter, codenamed P92. Is there something you can say about those rumors? There is a project. You know, I've always thought that sort of a text-based kind of information utility is just a really important thing to society. And

for whatever reason, I feel like Twitter has not lived up to what I would have thought its full potential should be. And I think that the current, you know, I think Elon thinks that right. And that's probably one of the reasons why you bought it. And I do know there are ways to consider alternative approaches to this. And one that I think is potentially interesting is this open and federated approach where you're seeing with Mastodon. I mean, you're seeing that a little bit with Blue Sky. And I think that it's possible that something that melds some of those ideas with the graph and identity system that people have already cultivated on Instagram could be a kind of very welcome contribution to that space. But I know we work on a lot of things all the time though, too. So I don't want to get ahead of myself. And we have projects that explore a lot of different things. And this is certainly one that I think could be interesting. So what's the release, the launch date of that again?

Or what's the official website? And we don't have that yet.

Oh, okay. And look, I mean, I don't know exactly how this is going to turn out. I mean, what I can say is, yeah, there's some people working on this. I think that there's something there that's interesting to explore. So if you look at, it'd be interesting to just ask this question and throw Twitter into the mix, that the landscape of social networks that is Facebook, that is Instagram, that is WhatsApp, and then think of a text-based social network.

When you look at that landscape, what are the interesting differences to you?

Why do we have these different flavors? And what are the needs, what are the use cases, what are the products? What is the aspect of them that create a fulfilling human experience and a connection between humans that is somehow distinct?

Well, I think text is very accessible for people to transmit ideas and to have back and forth exchanges. So it I think ends up being a good format for discussion, in a lot of ways, uniquely good. If you look at some of the other formats or other networks that have focused on one type of content like TikTok is obviously huge. And there are comments on TikTok. But I think the architecture of the service is very clearly that you have the video is the primary thing, and there's comments after that. But I think one of the unique pieces of having text-based content is that the comments can also be first class. And that makes it so that conversations can just filter and fork into all these different directions and in a way that can be super useful. There's a lot of things that are really awesome about the experience. It just always struck me, I always thought that Twitter should have a billion people using it, or whatever the thing is that basically ends up being in that space. And for whatever combination of reasons, again, these companies are complex organisms and it's very hard to diagnose this stuff from the outside. Why doesn't Twitter, why doesn't a text-based comment as a first citizen based social network have a billion users? Well, I just think it's hard to build these companies. So it's not that every idea automatically goes and gets a billion people. It's just that I think that that idea coupled with good execution should

get there. But I mean, look, we hit certain thresholds over time where we plateaued early on and it wasn't clear that we were ever going to reach 100 million people on Facebook. And then we got really good at dialing in internationalization and helping the service grow in different countries. And that was a whole competence that we needed to develop and helping people basically spread the service to their friends. That was one of the things, once we got very good at that, that was one of the things that made me feel like, hey, if Instagram joined us early on, then I felt like we could help grow that quickly and same with WhatsApp. And that's sort of been a core competence that we've developed and been able to execute on. And others have too, right? I mean, ByteDance obviously have done a very good job with TikTok and have reached more than a billion

people there. But it's certainly not automatic, right? I think you need a certain level of execution to basically get there. And I think for whatever reason, I think Twitter has this great idea and sort of magic in the service. But they just haven't kind of cracked that piece yet. And I think that that's made it so that you're seeing all these other things, whether it's Mastodon or Blue Sky, that I think are maybe just different cuts at the same thing. But I think through the last generation of social media overall, one of the interesting experiments that I think should get run at larger scale is what happens if there's somewhat more decentralized control. And if it's the stack is more open throughout. And I've just been pretty fascinated by that and seeing how that works. To some degree, end-to-end encryption on WhatsApp and as we bring it to other services provides an element of it because it pushes the service really out to the edges. The server part of this that we run for WhatsApp is relatively very thin compared to what we do on Facebook or Instagram. And much more of the complexity is how the apps kind of negotiate with each other to pass

information in a fully end-to-end encrypted way. But I don't know, I think that that is a good model. I think it puts more power in individuals' hands and there are a lot of benefits of it if you can make it happen. Again, this is all pretty speculative. I mean, I think that it's hard from the outside to know why anything does or doesn't work until you kind of take a run at it. And so I think it's kind of an interesting thing to experiment with, but I don't really know where this one's going to go. So since we were talking about Twitter, Elon Musk had what I think a few harsh words that I wish he didn't say. So let me ask, in the hope in the name of camaraderie, what do you think Elon is doing well with Twitter? And as a person who has run for a long time, you social networks, Facebook, Instagram, WhatsApp, what can he do better? What can he improve on that text-based social network? Gosh, it's always very difficult to offer specific critiques from the outside before you get into this because I think one thing that I've learned is that everyone has opinions on what you should do and running the company, you see a lot of specific nuances on things that are not apparent externally. And I often think that some of the discourse around us could be better if there is more kind of space for acknowledging that there's certain things that we're seeing internally that guide what we're doing. But I don't know. I mean, since you asked what is going well, you know, I do think that Elon led a push early on to make Twitter a lot leaner. And I think that that, you know, it's like you can agree or disagree with exactly all the tactics and how we did that. You know, obviously, you know, every leader has their own style for if they, you know, if you need to make dramatic changes for that, how you're going to execute it.

But a lot of the specific principles that he pushed on around basically trying to make the organization more technical around decreasing the distance between engineers at the company and him, like fewer layers of management. I think that those were generally good changes. And I'm also, I also think that it was probably good for the industry that he made those changes, because my sense is that there were a lot of other people who thought that those were good changes, but who may have been a little shy about doing them. And I think he, you know, and just in my conversations with other founders and how people have reacted to the things that we've done, you know, what I've heard from a lot of folks is, is just, hey, you know, when you, when someone like you, you know, when I wrote the letter outlining the organizational changes that I wanted to make back in March, and you know, when people see what Elon is doing, I think that that gives, you know, people the ability to think through how to shape their organizations in a way that, that, that, you know, hopefully can, can be good for the industry and make all these companies more productive over time. So something that that was one where I think he was quite ahead of a bunch of the other companies on and, you know, what he was doing there, you know, again, from the outside, very hard to know. It's like, okay, did he cut too much? Did he knock enough? Whatever. I don't think it's like my place to opine on that. And you asked for a, for a positive framing of the question of what, what do I, what do I admire? What do I think it went well? But I think that, like certainly his actions led me and I think a lot of other folks in the industry to think about, hey, are we, are we kind of doing this as much as we should? Like, can we, is, could we make our companies better by pushing on some of these same principles? Well, the two of you are on the top of the world in terms of leading the development of tech. And I wish there was more both way camaraderie and kindness, more love in the world, because love is the answer. But let me ask kind of a point of efficiency. You recently announced multiple stages of layoffs and meta. What are the most painful aspects of this process? Given for the individuals, the painful effects it has on those people's lives? Yeah, I mean, that's it. And that's it. I mean, it's a, and you basically have a significant number of people who, you know, this is just not the end of their time at meta that they or, or I, you know, would have hoped for when they joined the company. And, you know, I mean, running a company, people are, you know, constantly joining and leaving the company for different directions, but, but for different, different reasons. But, um, and layoffs are like uniquely challenging and tough in that you have a lot of people leaving for reasons that aren't connected to their own performance, or, you know, the, the culture not being a fit at that point, it's really just, it's a, it's a kind of strategy decision and sometimes financially required. Um, but not, not fully in our case, and especially on the changes that we made this year, a lot of it was more kind of culturally and strategically driven by this push where I wanted us to become a, a stronger technology company with a more of a focus on building a more technical and more of a focus on building higher quality products faster. And I just view the external world is quite volatile right now. And I wanted to make sure that we had a stable position to be able to continue investing in these long-term, ambitious projects that we have around, you know, continuing to push AI forward and continuing to push forward all the metaverse work. And in order to do that in light of the

pretty big thrash that we had seen over the last 18 months, you know, some of it, um, you know, macroeconomic induced, some of it specific, some of it competitively induced, some of it, um, just because of bad decisions, right, or things that we got wrong. Um, I don't know, I just, I decided that we needed to get to a point where we were a lot leaner and but look, I mean, but then, okay, it's, it's one thing to do that to like decide that at a high level. Then the question is, how do you execute that as compassionately as possible? And there's no good way. Um, there's no perfect way for sure. And it's, it's, it's gonna be tough no matter what, but I, you know, as, as a leadership team here, we've certainly spent a lot of time just thinking, okay, given that this is a thing that sucks, like what is the most compassionate way that we can do this? And, um, and that's what we've tried to do. And you mentioned there, there's an increased focus on, uh, engineering on tech. So the technology teams, tech focus teams on building products that, yeah, I mean, I wanted to, I want to empower engineers more, the people are building things, the tech, the technical teams. Um, part of that is making sure that the people are building things aren't just at like the leaf nodes of the organization. I don't want like, you know, eight levels of management and then the people actually doing the work. So we made changes to make it so that you have individual contributor engineers reporting at almost every level up the stack. Which I think is important because, you know, you're running a company, one of the big questions is, you know, latency of, of information that you get. You know, we talked about this a bit earlier in terms of kind of the joy of, of, of the feedback that you get doing something like jujitsu compared to running a long-term project. But I actually think part of the art of running a company is trying to constantly re-engineer it so that your feedback loops get shorter so you can learn faster. And part of the way that you do that is by, I kind of think that every, every layer that you have in the organization, um, means that information might not need to get reviewed before it, it, it goes to you. And I think, you know, making it so that the people doing the work are as close as possible to you as possible is, is, is pretty important. So there's that. I mean, I think over time, companies just build up very large support functions that are not doing the kind of core technical work. And those functions are very important, but I think having them in the right proportion is, is important. And if, um, if you, you try to do good work, but you don't have, you know, the right, you know, marketing team or, um, or the right legal advice, like you're going to, you know, make some pretty big blunders, but, um, but at the same time, if you have, you know, if, if you just like have too big of, of, of things and, and some of these support roles, then that might make it so that things are just move a lot. Um, maybe you're too conservative or you, you move a lot slower, um, uh, than, than, than you should otherwise. Those are just examples, but it's, um, but How do you find that balance? It's really tough. Yeah. No, I, but that's, it's a constant equilibrium that you're, that you're searching for. Yeah. How many managers to have? What are the pros and cons of managers? Well, I mean, I, I believe a lot in management. I think there are some people who think that it doesn't matter as much, but look, I mean, we have a lot of younger people at the company for them. This is their first job and, you know, people need to grow and learn in their career and like that. All that stuff is important, but here's one mathematical way to look at it.

Um, you know, at the beginning of this, we, um, I asked our, our people team was the average number of, of reports that a manager had. And I think it was, it was around three, maybe three to four, but closer to three. I was like, wow, like a manager can, you know, best practices that person can, can manage, you know, seven or eight people. Um, but there was a reason why it was closer to three. It was because we were growing so quickly, right? And when you're hiring so many people so quickly, then that means that you need managers who have capacity to onboard new people. Um, and also if you have a new manager, you may not want to have them have seven direct reports immediately because you want them to ramp up. But the thing is going forward, I don't want us to actually hire that many people that quickly, right? So I actually think we'll just do better work if we have more constraints and we're, um, you know, leaner as an organization. So in a world where we're not adding so many people as quickly, is it as valuable to have a lot of managers who have extra capacity waiting for new people? No, right? So, um, so now we can, we can sort of defragment the organization and get to a place where the average is closer to that seven or eight. Um, and it's, it's just ends up being a somewhat more kind of compact management structure, which, um, you know, decreases the latency on, on information going up and down the chain. And, um, and I think empowers people more, but I mean, that's, that's an example that I think it doesn't kind of undervalue the importance of management and, and the, um, kind of the personal growth or coaching that people need in order to do their jobs. Well, it's just, I think realistically, we're, we're just not going to hire as many people going forward. So I think that you need a different structure. This whole, this whole incredible hierarchy and network of humans that make up a company is fascinating. How do you hire great teams? How do you hire great now with the focus on engineering and technical teams? How do you hire great engineers and great members of technical teams? Well, you're asking how you select or how you attract them? Both, but select, I think, uh, I think attract is work on cool stuff and have a vision. I think that's right. And, and, and have a track record that people think you're actually going to be able to do it. Yeah. To me, the select is, seems like more of the art form, more of the tricky thing. How do you select the people that fit the culture and can get integrated the most effectively and so on. And maybe, especially when they're young, to see like, to see the magic through the, through the resume, through the paperwork and all this kind of stuff, to see that there's a special human there that would do like incredible work. So there are lots of different cuts on this question. I mean, I think when an organization has grown quickly, one of the big questions that teams face is, do I hire this person who's in front of me now because they seem good? Or do I hold out to get someone who's even better? And the heuristic that I always focused on for myself and my own kind of direct hiring that I think works. So when you, when you recurs it through the organization is that you should only hire someone to be on your team if you would be happy working for them in an alternate universe. And something that, that kind of works. And that's basically how I've tried to build my team. It's, you know, I'm not, I'm not in a rush to not be running the company. But I think in an alternate universe where one

of these other folks was running the company, I'd be happy to work for them. I feel like I'd learn from them. I respect their kind of general judgment. They're all very insightful. They have good values. And I think that that gives you some rubric for, you can apply that at every layer. And I think if you apply that at every layer in the organization, then you'll have a pretty strong organization. Okay, in an organization that's not growing as quickly, the questions might be a little different though. And there, you asked about young people specifically, like people out of college. And one of the things that we see is it's, it's a pretty basic lesson, but like, we have a much better sense of who the best people are, who have interned at the company for a couple of months, then by looking at them at kind of a resume or a short, or a short interview loop. I mean, obviously the in-person feel that you get from someone probably tells you more than the resume. And you can do some basic skills assessment. But a lot of the stuff really just is cultural. People thrive in different environments. And on different teams, even within a specific company, and it's like the people who come for even a short period of time over a summer, who do a great job here, you know that they're going to be great if they came and joined full time. And that's, you know, one of the reasons why we've invested so much in internship is basically it just, it's a very useful sorting function, both for us and for the people who want to try out the company. You mentioned in-person, what do you think about remote work, a topic that's been discussed extensively because of over the past few years, because of the pandemic? Yeah, I mean, I think it's, I mean, it's, it's a thing that's here to stay. But I think that there's, there's value in both, right? It's not, you know, I wouldn't want to run a fully remote company yet, at least. I think there's an asterisk on that, which is that, some of the other stuff you're working on, yeah. Yeah, exactly. It's like all the, all the, you know, metaverse work and the ability to be, to feel like you're truly present, no matter where you are. I think once you have that all dialed in, then we may, you know, one day reach a point where it really just doesn't matter as much where you are physically. But I don't know, today it, today it still does, right? So, yeah, for people who, there are all these people who have special skills and want to live in a place where we don't have an office, are we better off having them at the company? Absolutely, right? And are a lot of people who work at the company for several years and then, you know, build up the relationships internally and kind of have the trust and have a sense of how the company works. Can they go work remotely now if they want and still do it as effectively? And we've done all these studies that show it's like, okay, does that affect their performance? It does not. But, you know, for the new folks who are joining and for people who are earlier in their career and you don't need to learn how to solve certain problems and need to get ramped up on the culture, you know, when you're working through really complicated problems where you don't just want to sit in the, you don't just want the formal meeting, but you want to be able to like brainstorm when you're walking in the hallway together after the meeting. I don't know, it's like we just haven't replaced the kind of in-person dynamics there yet with anything remote yet. So, yeah, there's a magic to the in-person that we'll talk about this a little bit more, but I'm really excited by the possibilities in the next two years in virtual reality and mixed reality that are possible with high-resolution scans. I mean,

I, as a person who loves in-person interaction, like these podcasts in person, it would be incredible to achieve the level of realism I've gotten the chance to witness. But let me ask about that. Yeah. I got a chance to look at the Quest 3 headset and it is amazing. You've announced it. It's, you'll give some more details in the fall, maybe release in the fall. When is it getting released again? I forgot you mentioned it. We'll give more details at Connect, but it's coming this fall. Okay.

So, it's priced at \$4.99. What features are you most excited about there?

There are basically two big new things that we've added to Quest 3 over Quest 2. The first is high-resolution mixed reality. And the basic idea here is that you can think about virtual reality as you have the headset and all the pixels are virtual and you're basically like immersed in a different world. Mixed reality is where you see the physical world around you and you can place virtual objects in it, whether that's a screen to watch a movie or a projection of your virtual desktop, or you're playing a game where like zombies are coming out through the wall and you need to shoot them. Or, you know, we're, you know, we're playing Dungeons & Dragons or some board game and we just have a virtual version of the board in front of us while we're sitting here. All that's possible in mixed reality and I think that that is going to be the next big capability on top of virtual reality.

It has done so well. I have to say, as a person who experienced it today with zombies, having a full awareness of the environment and integrating that environment in the way they run at you while they try to kill you, it's just the mixed reality the pass through is really, really, really well done. And the fact that it's only \$500 is really, it's well done.

Thank you. I mean, I'm super excited about it. I mean, we put a lot of work into making the device both as good as possible and as affordable as possible because a big part of our mission and ethos here is we want people to be able to connect with each other. We want to reach and we want to serve a lot of people. We want to bring this technology to everyone. So, we're not just trying to serve like an elite, a wealthy crowd. We really want this to be accessible. So, that is in a lot of ways an extremely hard technical problem because, you know, we don't just have the ability to put an unlimited amount of hardware and that's we needed to basically deliver something that works really well but in an affordable package. And we started with Quest Pro last year. It was \$1,500 and now we've lowered the price to \$1,000. But in a lot of ways, the mixed reality in Quest 3 is an even better and more advanced level than what we were able to deliver in Quest Pro. So, I'm really proud of where we are with Quest 3 on that. It's going to work with all of the virtual reality titles and everything that existed there. So, people who want to play fully immersive games, social experiences, fitness, all that stuff will work. But now you'll also get mixed reality too, which I think people really like because it's sometimes you want to be super immersed in a game. But a lot of the time, especially when you're moving around, if you're active, like you're doing some fitness experience, you know, let's say you're like doing boxing or something. It's like you kind of want to be able to see the room around you. So, that way you know that like I'm not going to punch a lamp or something

like that. And I don't know if you got to play with this experience, but we basically have the and it's just sort of like a fun little demo that we put together. But it's like you just, we're like in a conference room or you're living room and you have the guy there and you're boxing

him and you're fighting him and it's like. Oh, the other people are there too. I got a chance to do that. Yeah. And all the people are there. It's like that guy is right there. Yeah, just like it's right in the room. And the other human, the path that you're seeing them also, they can cheer you on, they can make fun of you if there are anything like friends of mine. And then just it, yeah, it's really, it's a really compelling experience. And VR is really interesting too, but this is something else almost. This is, this becomes integrated into your life, into your world. Yeah. And it, so I think it's a completely new capability that will unlock a lot of different content. And I think it'll also just make the experience more comfortable for a set of people who didn't want to have only fully immersive experiences. I think if you want experiences where you're grounded in, you know, your living room in the physical world around you, now you'll be able to have that too. And I think that that's pretty exciting.

I really liked how it added windows to a room with no windows. Yeah. Me as a person.

Did you see the aquarium one where you could see the sharks swim up or was that just a zombie one?

Just a zombie one, but it's still off the track. You don't necessarily want windows added to your living room where zombies come out of, but yeah, so the context of that game, it's, yeah, yeah.

I enjoyed it because you could see the nature outside. And me as a person that doesn't have windows, it's just nice to have nature. Yeah. Well, even if it's a mixed reality setting.

I know it's a zombie game, but there's a zen nature, zen aspect to being able to look outside and alter your environment as you know it. Yeah. And there will probably be better, more zen ways to do that than the zombie game you're describing, but you're right that the basic idea of sort of having your physical environment on pass through, but then being able to bring in different elements. I mean, I think it's going to be super powerful. And in some ways, I think that these are mixed realities, also a predecessor to eventually we will get AR glasses that are not kind of the goggles form factor of the current generation of headsets that people are making. But I think a lot of the experiences that developers are making for mixed reality of basically you just have a kind of a hologram that you're putting in the world will hopefully apply once we get the AR glasses too. Now that's got its own whole set of challenges and it's... Well, the headsets are already smaller than the previous version.

Oh yeah, it's 40% thinner. And the other thing that I think is good about it, yeah,

so mixed reality was the first big thing. The second is it's just a great VR headset.

It's got 2x the graphics processing power, 40% sharper screens, 40% thinner, more comfortable, better strap architecture, all this stuff that if you liked Quest 2, I think that this is just going to be... It's like all the content that you might have played in Quest 2 is just going to be sharper automatically and look better in this. So I think people are really going to like it.

Yeah, so this fall. This fall, I have to ask, Apple just announced a mixed reality headset called Vision Pro for \$3,500 available in early 2024. What do you think about this headset? Well, I saw the materials when they launched. I haven't gotten a chance to play with it yet. So kind of take everything with a grain of salt, but a few high-level thoughts. First, I do think that this is a certain level of validation for the category where we were the primary folks out there before saying, hey, I think that this virtual reality, augmented reality, mixed reality, this is going to be a big part of the next computing platform.

I think having Apple come in and share that vision

will make a lot of people who are fans of their products really consider that. And then, of course, the \$3,500 price, on the one hand, I get it with all the stuff that they're trying to pack in there. On the other hand, a lot of people aren't going to find that to be affordable. So I think that there's a chance that them coming in actually increases demand for the overall space and that Quest 3 is actually the primary beneficiary of that because a lot of the people who might say, hey, I'm going to give another consideration to this or now I understand maybe what mixed reality is more and Quest 3 is the best one on the market that I can afford. And it's great also. In our own way, I think there are a lot of features that we have where we're leading on. So I think that that could be quite good. And then obviously over time, the companies are just focused on somewhat different things. Apple has always focused on building really kind of high-end things, whereas our focus has been on, we have a more democratic ethos. We want to build things that are accessible to a wider number of people. We've sold tens of millions of Quest devices. My understanding, just based on rumors, I don't have any special knowledge on this, is that Apple is building about one million of their device. So just in terms of what you kind of expect in terms of sales numbers, I just think that Quest is going to be the primary thing that people in the market will continue using for their foreseeable future. And then obviously over the long term, it's up to the companies to see how well we've executed the different things that we're doing. But we kind of come at it from different places. We're very focused on social interaction, communication, being more active. So there's fitness, there's gaming, there are those things. Whereas I think a lot of the use cases that you saw in Apple's launch material were more around people sitting, people looking at screens, which are great. I think that you will replace your laptop over time with a headset. But I think in terms of kind of how the different use cases that the companies are going after, they're a bit different for where we are right now. Yeah. So gaming wasn't a big part of the presentation, which is an interesting, it feels like mixed reality gaming is such a big part of that. It was interesting to see it missing in the presentation. Well, I mean, look, there are certain design trade-offs in this where, you know, they only made this point about not wanting to have controllers, which on the one hand, there's a certain elegance about just being able to navigate the system with eye gaze and hand tracking. And by the way, you'll be able to just navigate quest with your hands too, if that's what you want. One of the things I should mention is that the capability from the cameras to, with computer vision to detect certain aspects of the hand, allowing you to have a controller that doesn't have that ring thing. Yeah. The hand tracking in Quest 3 and the controller tracking is a big step up from the last generation. And one of the demos that we have is basically an MR experience teaching you how to play piano, where it basically highlights the notes that you need to play and it's like, it's just all, it's hands, it's no controllers. But I think if you care about gaming, having a controller allows you to have a more tactile and allows you to capture fine motor movement much more precisely than what you can do with hands without something that you're touching. So again, I think there are certain questions which are just around what use cases are you optimizing for. I think if you want to play games, then I think that that, then I think you want, you want to design the system in a different way and we're more focused on kind of social experiences, entertainment experiences. Whereas if what you want is to make sure that the text that you read on a screen is as crisp as

possible, then you need to make the design and cost trade-offs that they made that lead you to making a \$3,500 device. So I think that there is a use case for that for sure, but I just think that they're, they've, the company is, we've basically made different design trade-offs to get to the use cases that we're trying to serve. There's a lot of other stuff I'd love to talk to you about, about the metaverse, especially the Kodak avatar, which I've gotten to experience a lot of different variations of recently that I'm really, really excited about. I'm excited to talk about that too. I'll, I'll have to wait a little bit because,

well, I think there's a lot more to show off in that regard. But let me step back to AI.

I think we've mentioned it a little bit, but I'd like to linger on this question that folks like Eliazer Yudkowsky has a worry about and others of the existential, the serious threats of AI that have been reinvigorated now with the rapid developments of AI systems. Do you worry about the existential risks of AI as Eliazer does about the alignment problem, about this getting out of hand? Anytime where there's a number of serious people who are raising a concern that is that existential about something that you're involved with, I think you have to think about it, right? So I've spent quite a bit of time thinking about it from that perspective.

The thing that I, where I basically have come out on this for now is I do think that there are, over time, I think that we need to think about this even more as we, as we approach something that, you know, could be closer to superintelligence. I just think it's pretty clear to anyone working on these projects today that we're not there. And one of my concerns is that we spent a fair amount of time on this before, but there are more, I don't know if mundane is the right word, but there's like concerns that already exist right about people using AI tools to do harmful things of the type that we're already aware, whether, you know, we talked about fraud or scams or different things like that. And that's going to be a pretty big set of challenges that the companies working on this are going to need to grapple with, regardless of whether there is an existential concern as well at some point down the road. So I do worry that to some degree you can, people can get a little too focused on, on some of the tail risk and then not do as good of a job as we need to on the things that you can be almost certain are going to come down the pipe as real risks that kind of manifest themselves in the near term. So for me, I've spent most of my time on that once I kind of made the realization that the size of models that we're talking about now in terms of what we're building are just quite far from the superintelligence type concerns that that people raise. But I think once we get a couple steps closer to that,

I know as we do get closer, I think that those, you know, there are going to be some novel risks and issues about how we make sure that the systems are safe for sure. I guess here, just to take the conversation in a somewhat different direction, I think in some of these debates around safety, I think the concepts of intelligence and autonomy or like the, the, the being of the thing, you know, as an analogy, they get kind of conflated together. And I think it very well could be the case that you can make something and scale intelligence quite far, but that, that may not manifest the safety concerns that people are saying in the sense that I mean, just if you, if you look at human biology, it's like, all right, we have our neocortexes where all the thinking happens, right? And it's, but, but it's not really calling the shots at the end of the day. We have a much more, you know, primitive old brain structure for which our neocortex, which is this powerful machinery is basically just a kind of prediction and reasoning engine to help it kind of like our very simple brain decide how to plan and do what it needs to do

in order to achieve these like very kind of basic impulses. And I think that you can think about some of the development of intelligence along the same lines where just like our neocortex doesn't have free will or autonomy, we might develop these wildly intelligent systems that are much more intelligent than our neocortex have much more capacity, but are, you know, in the same way that our neocortex is sort of subservient and is used as a tool by our, our kind of simple impulse brain. It's, you know, I think that it's not out of the question that very intelligent systems that, that have the capacity to think we'll, we'll kind of act as that is sort of an extension of, of the neocortex doing that. So I think my own view is that where we really need to be careful is on the development of autonomy and how we think about that. Because it's actually the case that relatively simple and unintelligent things that have runaway autonomy and just spread themselves or, you know, it's like, we have a word for that. It's a virus, right? It's, I mean, like it's, can be simple computer code that is not particularly intelligent, but just spreads itself and does a lot of harm, you know, biologically or computer. And I just think that these are somewhat separable things. And a lot of what I think we need to develop when people talk about safety and responsibility is really the governance on the autonomy that can be given to, to systems. And to me, if, you know, if I were, you know, a policymaker is, or thinking about this, I would really want to think about that distinction between these, where I think building intelligent systems will be, can create a huge advance in terms of people's quality of life and productivity growth in the economy. But it's the autonomy part of this, that I think we really need to make progress on how to govern these things responsibly before we build the capacity for them to make a lot of decisions on their own or, or give them goals or things like that. And I know that that's a research problem, but I do think that to some degree, these are, are somewhat, are somewhat separable things. I love the distinction between intelligence and autonomy and, and the metaphor within New York Cortex. Let me ask about power. So building superintelligence systems, even if it's not in the near term, I think meta as is one of the few companies, if not the main company that will develop the superintelligence system. And you are a man who's at the head of this company, building AGI might make you the most powerful man in the world. Do you worry that that power will corrupt you? What a question. I mean, look, I think realistically, this gets back to the open source things that we talked about before, which is, I don't think that the world will be best served by any small number of organizations having this without it being something that is more broadly available. And I think if you look through history, it's when there are these sort of like unipolar advances and things that, and like power imbalances that they're, they're, they're doing to being kind of weird situations. So this is one of the reasons why I think open sources is, is generally the right approach. And, you know, I think it's a, it's a categorically different question today when we're not close to superintelligence. I think that there's a good chance that even once we get closer to superintelligence, open sourcing remains the right approach, even though I think at that point it's a somewhat different debate. But I think part of that is that that is, you know, I think one of the best ways to ensure that the system is as secure and safe as possible, because it's not just about a lot of people having access to it. It's the scrutiny that, that, that kind of comes with being, with building an open source system, right? I think that this is a pretty widely accepted thing about open sources that you have the code out there so

anyone can see the vulnerabilities. Anyone can, can kind of mess with it in different ways. People can spin off their own projects and experiment in a ton of different ways. And the net result of all of that is that the systems just get hardened and get to be a lot safer and more secure. So I think that there's a chance that that ends up being the way that this goes to a pretty good chance and that having this be open both leads to a healthier development of the technology and also leads to a more balanced distribution of the technology in a way that, that strike me as good values to aspire to. So to you the risks, there's risks to open sourcing, but the benefits outweigh the risks. At the two, it's interesting. I think the way you put it, you put it well that there's a different discussion now than when we get closer to the, to development of super intelligence of, of the benefits and risks of open sourcing. Yeah. And to be clear, I feel quite confident in the assessment that open sourcing models now is net positive. I think there's a good argument that in the future it will be two, even as you get closer to super intelligence. But I've not, I've certainly have not decided on that yet. And I think that it becomes a somewhat more complex set of questions that I think people will have time to debate and will also be informed by what happens between now and then and to make those decisions. We don't have to necessarily just debate that in theory right now. What year do you think we'll have a super intelligence? I don't know. I mean, that's pure speculation. I think it's, I think it's very clear to take a step back that we had a big breakthrough in the last year, right? Where the, the LLMs and diffusion models basically reached a, a scale where they're able to do some, some pretty interesting things. And then I think the question is what happens from here. And just to paint the two extremes on the, on one side, it's like, okay, we just had one breakthrough. If we just have like another breakthrough like that, or maybe two, then we can have something that's truly crazy, right? And, and is like, is just like so much more advanced. And, and on that side of the argument, it's like, okay, well, maybe we're, maybe we're only a couple of big steps away from, from, from, from reaching something that looks more like general intelligence. Okay, that's one, that's one side of the argument. And the other side, which is what we've historically seen a lot more is that a breakthrough leads to you know, in that, in that Gartner hype cycle, there's like the hype, and then there's the trough of disillusionment after when like people think that there's a chance that, hey, okay, there's a big breakthrough, maybe we're about to get another big breakthrough. And it's like, actually, you're not about to get another breakthrough, you're, maybe you're actually just gonna have to sit with this one for a while. And, and you know, it could be, could be five years, it could be 10 years, it could be 15 years until you figure out the, the kind of the next big thing that needs to get figured out. And, but I think that the fact that we just had this breakthrough, sort of makes it's that we're at a point of almost a very wide error bars on what happens next. I think the traditional technical view, or the, like looking at the industry would suggest that we're not just going to stack in a like breakthrough on top of breakthrough on top of breakthrough, like every six months or something right now. I think it will, I'm guessing, I would guess that it will, that it will take somewhat longer in between these, but I don't know. I tend to be pretty optimistic about breakthroughs too. So I mean, so I think if you, if you, if you normalized for, for my normal optimism, then, then maybe it would be even, even slower than what I'm saying. But,

but even within that, like I'm not even opining on the question of how many breakthroughs are required to get to general intelligence because no one knows. But this particular breakthrough was so

such a small step that resulted in such a big leap in performance as experienced by human beings, that it makes you think, wow, are we, as we stumble across this very open world of research, will we stumble across another thing that will have a giant leap in performance?

And also we don't know exactly at which stage is it really going to be impressive, because it feels like it's really encroaching on impressive levels of intelligence.

You still didn't answer the question of what year we're going to have superintelligence. I'd like to hold you to that. No, I'm just kidding. But is there something you could say about the timeline as you think about the development of AGI, superintelligence systems?

Sure. So I, I still don't think I have any particular insight on when like a singular AI system that is a general intelligence will get created. But I think the one thing that most people in the discourse that I've seen about this haven't really grappled with is that we do seem to have organizations and structures in the world that exhibit greater than human intelligence already.

So one example is a company. It acts as an entity. It has a singular brand. Obviously, it's a collection of people. But I certainly hope that meta with tens of thousands of people makes smarter decisions than one person. But I think that that would be pretty bad if it didn't. But another example that I think is even more removed from kind of the way we think about like the personification of, of, of intelligence, which is often implied in some of these questions is think about something like the stock market, right? The stock market is, you know, takes inputs. It's a distributed system. It's like the cybernetic organism that, you know, probably millions of people around the world are basically voting every day by choosing what to invest in. But it's basically this, this organism or, or structure that is smarter than any individual that we use to allocate capital as efficiently as possible around the world. And I do think that this notion that there are already these cybernetic systems that are either melding the intelligence of multiple people together or melding the intelligence of multiple people and technology together to form something which is dramatically more intelligent than any individual

on the, in the world is something that seems to exist and that we seem to be able to harness in a productive way for our society as long as we basically build these structures and balance with each other. So I don't know. I mean, that, that at least gives me hope that as we advance the technology, and I don't know how long exactly it's going to be, but you asked, when is this going to exist? I think to some degree, we already have many organizations in the world that are smarter than a single human. And, and that seems to be something that is generally productive in advancing humanity. And somehow the individual AI systems empower the individual humans and the

interaction between those humans to make that collective intelligence machinery that you're referring to smarter. So it's not like AI is becoming super intelligent. It's just becoming the, the engine that's making the collective intelligence is primarily human, more intelligent.

Yeah, it's educating the humans better. It's making them better informed. It's making it more efficient for them to communicate effectively and debate ideas. And through that process, just making the whole collective intelligence more and more and more intelligent,

maybe faster than the individual AI systems that are trained on human data anyway, are becoming. Maybe the collective intelligence of the human species might outpace the development of AI. I think there's a balance in here because I mean, if, if like, you know, if a lot of the input that, that the systems are being trained on is basically coming from feedback from people, then a lot of the development does need to happen in human time, right? It's, it's not like a machine will just be able to go learn all the stuff about, about how people think about stuff. There's, there's a cycle to how this needs to work. This is an exciting world we're living in and that you're at the forefront of developing. One of the ways you keep yourself humble, like we mentioned with Jiu Jitsu, is doing some really difficult challenges, mental and physical. One of those you've done very recently is the Murph challenge. And you got a really good time. It's a hundred pull-ups, 200 push-ups, 300 squats and a mile before and a mile run after. You got under 40 minutes on that. What was the hardest part? I think a lot of people were very impressed. It's very impressive time. How crazy are you? It wasn't my best time, but, but I, anything under 40 minutes, I'm happy with. It wasn't your best time? No, I think, I think I've done it a little faster before, but not much. I mean, it's, and of my friends, I did not win on Memorial Day. One of my friends did it actually several minutes faster than me. But just to clear up one thing that I think was, I saw a bunch of questions about this on the internet. There are multiple ways to do, to do the Murph challenge. There's a kind of partitioned mode where you do sets of pull-ups, push-ups and squats together. And then there's unpartitioned where you do the hundred pull-ups and then the 200 push-ups and then the 300 squats in cereal. And obviously, if you're, you know, if you're doing them unpartitioned, then, you know, it takes longer to get through the hundred pull-ups because you, you know, anytime you're resting in between the pull-ups, you're not also doing push-ups and, and squats. So, so yeah, so my, my, I'm sure my unpartition time would be, would be quite a bit slower, but, but no, I think at the end of this, I don't know, first of all, I think it's a good way to honor Memorial Day, right? It's, you know, it's this Lieutenant Murphy, basically, this is one of, this was one of his favorite exercises and I just try to do it on, on Memorial Day each year and it's a good workout. I got my older daughters to do it with me this time. They, my oldest daughter wants a weight vest because she sees me doing it with a weight vest. I don't know if a seven-year-old should be using a weight vest to do pull-ups, but, but, um... The difficult question a parent must ask themselves, yes. I was like, maybe I can make you a very light weight vest, but, but I, I didn't think it was good for this. So, she, she basically did a quarter, Murph, so she ran a quarter mile and then did, you know, 25 pull-ups, 50 push-ups and, and 75 air squats, then ran another quarter mile and like, in 15 minutes, which I was pretty impressed by, um, and, and my, my five-year-old too. So, I, so I was excited about that and I, I'm glad that I'm teaching them kind of the value of, kind of physicality, right? I think a, a good day for Max, my daughter, is when she gets to like, go to the gym with me and cranks out a bunch of pull-ups and I, I, I love that about her. I mean, I think it's, it's like, good. She's, you know, um, hopefully I'm teaching her some good lessons, but I mean, the, the broader question here is, um, given how busy you are, given how much stuff you have going on in your life, uh, what's, um, what's like the perfect exercise regimen for you to, uh, to keep yourself happy, to, uh, keep yourself productive in your main line of work? Yeah. So, I mean, I've, right now, I'm focused most of my workouts on, on fighting. So, so Jiu-Jitsu

and MMA, um, but I don't know. I mean, maybe if you're a professional, you can do that every day. I can't. I just get, you know, it's too many, too many bruises and things that you need to recover from. So, so I do that, you know, three to four times a week and then, um, and then the other day is, um, I just try to do a mix of things, like just cardio conditioning, strength building, mobility. Um, so you try to do something physical every day? Yeah, I try to, unless I'm just so tired that I just need to, need to relax, but then I'll still try to like go for a walk or something. I mean, even here, um, I don't know. I mean, have you, have you been on the roof here yet? No. We'll go on the roof after this. I heard things.

But it's like, I'm, we, we designed this, this building and I put a park on the, on the roof. So that way, that's like my, my meetings when I'm just doing kind of a one-on-one or talking to a couple of people, I'm, I, I have a very hard time just sitting. I feel like you get super stiff. It like feels really bad. Um, but I don't know. Being physical is very important to me. I think it's, um, I do not believe this gets to the question about AI. I don't think that a being is just a mind. Um, and I think we're, we're kind of meant to do things and like physically and, and a lot of the sensations that we feel are, um, are, are connected to that. And I think that that's a lot of what makes you a human is, is basically, you know, having those, having, you know, that set of sensations and experiences around that coupled with a mind to reason about them. Um, but I don't

know. I think it's important for balance to, to kind of get out, challenge yourself in different ways, learn different skills, clear your mind. Do you think AI, in order to become super intelligent, need you, I should have a body? It depends on, on what the goal is. I think that there's this assumption in that question that intelligence, intelligence should be kind of person-like. Whereas, you know, as we were just talking about, um, you can have these greater than single human

intelligent organisms, like the stock market, which obviously do not have bodies and do not speak a language, right? And like, you know, and, and just kind of have their own system. Um, but so I don't know, my guess is, um, there will be limits to what a system that is purely an intelligence can understand about the human condition without having the same, not just senses, but like our, our body's changes we get older, right? And, and we kind of evolve and like let those very subtle physical changes just drive a lot of social patterns and behavior around like when you choose to have kids, right? Like just like all these, you know, that's not even subtle. That's a major one, right? But like, um, you know, how you design things around the house. Um, so yeah, I mean, I think, I think it would, if the goal is to understand people as much as possible, I think, I think that that's trying to model those sensations is probably somewhat important, but I think that there's a lot of value that can be created by having intelligence, even that, that is, that is separate from that is a separate thing.

So one of the features of being human is that we're mortal. We die. We've talked about AI a lot, about potentially replicas of ourselves. Do you think there will be AI replicas of you and me that persist long after we're gone, that family and loved ones can talk to?

I think we'll have the capacity to do something like that. And I think one of the big questions that we've had to struggle with in the context of social networks is who gets to make that.

Um, and you know, my answer to that, you know, in the context of the work that we're doing is that that should be your choice, right? I don't think anyone should be able to choose to make

a lex spot that people can, can choose to talk to and get to train that. And we've, we've kind of, we have this precedent of making some of these calls where, I mean, someone can create a page for a lex fan club, but you can't create a page and say that you're lex, right? So I think that this, similarly, I think, I mean, maybe, you know, someone maybe can make a, should be able to make an AI

that's, that's a lex admirer that someone can talk to. But I think it should ultimately be your call whether there is a lex AI. Well, I'm open sourcing the lex.

So you're a man of faith. What, what role has faith played in your life and your understanding of the world and your understanding of your own life and your understanding of your work and how to, your work impacts the world. Yeah, I think that there's a few different parts of this that are relevant. There's sort of a philosophical part and there's a cultural part. And one of the most basic lessons is right at the beginning of Genesis where it's like God creates the earth and creates people and creates people in God's image. And there's the question of, you know, what does that mean? And all the only context that you have about God at that point in the Old Testament is that he's, God has created things. So I always thought that like one of the interesting lessons from that is that there's a virtue in creating things that is like whether it's artistic or whether you're building things that are functionally useful for other people.

I think that that by itself is a good. And I, that kind of drives a lot of how I think about morality and my personal philosophy around like what, what is a good life? Right? I think it's one where you're, you know, helping the people around you and you're being a kind of positive creative force in the world that is helping to, you know, bring new things into the world, whether they're, you know, amazing other people, kids, or, or just leading to the creation of different things that that wouldn't have been possible otherwise. And so that's a value for me that that matters deeply. And I just, I mean, I just love, you know, spending time with the kids and seeing that they sort of, you know, trying to impart this value to them. And it's like, I mean, nothing makes me happier than like when I come home from work and, you know, I see like my, my daughter's like building Legos on the table or something. It's like, all right, I did that when I was a kid, right? So many other people were doing this. And like, I hope you don't lose that spirit where when you, you kind of grow up and you want to just continue building different things no matter what it is. To me, that's a lot of what matters.

That's the philosophical piece. I think the cultural piece is just about community and values. And that part of things I think has just become a lot more important to me since I've had kids. You know, it's almost autopilot when you're a kid. You're in the kind of getting imparted two phase of your life. But, and I didn't really think about religion that much for a while. You know, I was in college, you know, before I, before I had kids. And then I think having kids has this way of really making you think about what traditions you want to impart. And, and how you want to celebrate and like what, what balance you want in your life. And I mean, a bunch of the questions that you've asked and a bunch of the things that we're talking about. Just the irony of the curtains coming down as we're talking about mortality.

Once again, same as last time. This is just, just that the universe works. And we are definitely living in a simulation. But go ahead. Community, tradition and the values, the faith and religion is still. A lot of the topics that we've talked about today are around how do you, how do you balance, you know, whether it's running a company or, or different responsibilities with

[Transcript] Lex Fridman Podcast / #383 - Mark Zuckerberg: Future of AI at Meta, Facebook, Instagram, and WhatsApp

this. Yeah. How do you, how do you kind of balance that? And I always also just think that's very grounding to just believe that there is something that is much bigger than you that is guiding things. That amongst other things gives, gives you a bit of humility.

As you pursue that spirit of creating that you, you spoke to creating beauty in the world.

And as Dostoevsky said, beauty will save the world. Mark, I'm a huge fan of yours.

Honored to be able to call you a friend and I am looking forward to

both kicking your ass and you kicking my ass on the mat tomorrow in Jiu Jitsu.

This, this incredible sport and art that we both, we both participate in. Thank you so much for talking to you. Thank you for everything you're doing in so many exciting realms of technology and human life. I can't wait to talk to you again in the metaverse. Thank you.

Thanks for listening to this conversation with Mark Zuckerberg. To support this podcast, please check out our sponsors in the description. And now let me leave you with some words from Isaac Asimov. It is change, continuing change, inevitable change that is the dominant factor in society today. No sensible decision can be made any longer without taking into account not only the world as it is, but the world as it will be. Thank you for listening and hope to see you next time.