

THE BROOKINGS INSTITUTION

WEBINAR

SHIFTING GEOPOLITICS IN THE AGE OF AI: A CONVERSATION WITH SAM ALTMAN

TUESDAY, MAY 07, 2024

MODERATORS:

MICHAEL E. O'HANLON

Senior Fellow and Director, Strobe Talbott Center for Security, Strategy, and Technology, Brookings

VALERIE WIRTSCHAFTER

Fellow, Foreign Policy, Artificial Intelligence and Emerging Technology Initiative, Brookings

DISCUSSANT:

SAM ALTMAN

Chief Executive Officer, OpenAI

* * * * *

O'HANLON: Greetings everyone, and welcome to this conversation with Sam Altman today I'm Mike O'Hanlon with the Foreign Policy program at Brookings and my colleague, Valerie Wirtschafter, is also here today. She will be interviewing Sam and talking about issues with disinformation, but also election protection, information protection and elections here at home and around the world. 2024 being an extremely important year, as we all know, for people going to vote in the United States and elsewhere. But at the same time, we're concerned about disinformation and new tools that the potential enemies of democracy or those who otherwise wish to interfere with proper elections, may try to employ whether here or abroad. Valerie is a senior fellow or, excuse me, a visiting fellow, but also a foreign policy scholar and our central person at Brookings with artificial intelligence and emerging technology an important priority of both our current president, Ceci Rouse and her predecessor, John Allen. So I will defer to her to have the conversation with Sam first and then come in a little bit later when Sam and I will discuss issues with war and peace and international security. Before going to some of your questions, just a brief word about Sam Altman. We all have been impressed and just wowed by what he's accomplished in his still young and yet extremely, impressive career already. He is now, for five years the CEO of OpenAI, a company that he helped create almost a decade ago. And that has taken the world by storm with its various remarkable innovations in artificial intelligence, not least ChatGPT, but others as well that he and Valerie will be discussing very soon. Sam hails from the great state of Missouri after having been born in Chicago, took an early interest in computer science, studied that at Stanford for a couple of years before he decided to go straight into the, innovation world, and so left college to form Loopt and then has been involved in, various startups over the years. And Y Combinator, of course, a very important one along the way where he helped many other startups get going. And now, OpenAI is anything but a startup. It's in many ways on the cusp of being one of the most important technological developments in the history of the human race. And so we're thrilled, Sam, to have you here. Thanks very much for joining us. And without further ado, I'll give the baton to my colleague Valerie to start the conversation.

WIRTSCHAFTER: Thanks so much, Mike, and thank you as well, Sam, for joining us today. So we're going to dive right in because I know we have a lot to discuss, and I'll try to weave in some questions we've already received from attendees. We're also going to try and incorporate at the end of our time, any additional questions that come in to events@brookings.edu. Now, of course, we've seen a lot of concern about AI generated content driving myths and disinformation around elections worldwide, much of which has not yet borne out fully in practice. However, a recent Microsoft Threat Intelligence report emphasized China's experimentation with AI driven disinformation campaigns, and the intelligence community highlighted

advances in AI as a new tool for foreign malign influence, particularly around elections. I think OpenAI also recently took action against five state affiliated malicious actors who were using ChatGPT for things like coding, support and translation. So I'm wondering, to start, how prevalent do you think this type of activity is across generative AI tools?

ALTMAN: As you pointed out, we have not seen the predicted flood of misinformation yet. We still might. But I think that there has been a lot of good work in the space. And also the platforms that distribute content online have been, I think, taking disinformation and misinformation seriously for quite some time now. And the detection and response has gotten better. What I'm worried about is not sort of more of the same, which again, I think we've built up technological defenses for and also societal antibodies. But the new stuff that may only be possible with AI. So the sort of the sophisticated one on one, 1 to 1 persuasion that you just couldn't do before. That may be possible with AI, I think, more of the conversation and more of the monitoring should be focused there.

WIRTSCHAFTER: Now kind of from a technical perspective, how does OpenAI just distinguish regular activity from malign usage? Is it primarily based, and I saw with those five, five actors that were part of the initial action. It was based on accounts. So is it primarily based on information, in the accounts, or is there another kind of strategy that you put in play?

ALTMAN: I mean, as many strategies as anyone that anyone who does this will say, you know, you get the best from looking at as many different signals as you possibly can. And then a really good investigations team.

WIRTSCHAFTER: I saw OpenAI shared an update today about its work in this area, with a focus on a new classifier for Dall E Three. A greater role in content's provenance efforts as part of the C2PA steering committee. Can you speak a little bit about what's new here?

ALTMAN: We, I mean, the, you know, an improved classifier based off of what we've learned and how people are attempting to misuse the tools, I think it's always good to do. But but I don't think about this as like, you know, yesterday we didn't do stuff and now we're doing more, it's that we do things continuously and, you know, periodically we try to provide an update. But if you look at how far our understanding of the tools and how to monitor the tools, has and how to align the tools has come, and also just the volume of how

much the world is using these tools. We're like every month is pretty different for us. So the kinds of threats to monitor for the way people are trying to misuse this, as I mentioned, this particular interest in looking for the customized one on one thing and any signals that that's being used, working with the platforms that are distributing this to get signal back there, like I view that is sort of all part of the package.

WIRTSCHAFTER: So how effective do you think? And I saw there were some limitations to the classifier that you all are putting out. But how effective do you think broadly the technical classifiers are? Are they going to kind of keep pace with the quality of content generation? And what more can be done in this area?

ALTMAN: One of the things about the field is we're still so early that the science is still advancing rapidly, and sometimes in ways we don't predict with what we know right now and how things seem to be evolving. I feel reasonably good about our ability to kind of stay abreast. But if the technology that we or others develop were to go in a very different direction, I think we'd have to reconsider that and think about new strategies.

WIRTSCHAFTER: Who should be thinking about these strategies? Is it the. Is it OpenAI's? Is it the companies developing the generative content or is it other other actors in this space?

ALTMAN: I think all of the above. I think that when this has worked really well, it is the technology developers, the application providers, the platforms, the government, civil society. I think it works the best when it's like a, a pretty high bandwidth and large conversation.

WIRTSCHAFTER: It's great. So we've also seen, some really kind of impressive new generative tools from OpenAI, Sora for video generation, Voice Engine for voice cloning. Of course there have been the public teasers, but they're not yet available for usage. So I'm curious about how OpenAI is thinking about the release of these tools, particularly given. And I would say audio is particularly challenging from both just a human association with audio and the intimacy around audio. But how is OpenAI thinking about the release of these tools, given these information integrity concerns? And I think maybe more broadly, to wrap that in a bigger question, what has changed about how OpenAI thinks about test products for safety? All the above. Since the launch launch of ChatGPT.

ALTMAN: Yeah. You know, for a long time we believed in iterative deployment, which is you don't want to go build something up with no feedback from society, no ability to observe how the world is using it, and also no time for society and the technology to co-evolved and to gradually get used to it. You also don't want to release it too late, especially when I think other people are going to release it, maybe with less of the care and thoughtfulness we'd hoped for anyway. But you don't want to release it too early either. And so we've had a playbook where we, you know, might preview a technology first and then release it to a few thousand users and then maybe a few million users and then more broadly and kind of have gates along the way in time to study that. Over time, a lot of things have changed. We've added, external red teamers before release and more recently that we've also added audits. And I think there will be new things to add as we, as we go forward. We'll do more, for example, with, government testing for our upcoming major release that we've done for, previous ones. And I, you know, for the next one after that, I'm sure there will be new things too. But this idea of preview the technology engage with external testers, release it to a small number of users, then a few more, and then eventually broadly. That is generally worked well for us.

WIRTSCHAFTER: That's great. Yeah. I mean, I think, you know, especially in considering the way that different tools are involving the different types of generated content. There might be different considerations for, an image detector versus an audio versus, video generation. And I think that, you know, having that type of iterative style is also really important, especially as these tools become, I think, quite, quite widespread. Of course, OpenAI, though, is not the only player in this space. And collaboration, particularly across the private sector, remains critical. As you said, of course, private. Some actors are more careful than others. But to this end, we saw a lot of the leading tech companies in the AI space sign an agreement with a really, you know, easy to pronounce title, the Tech Accord to Combat Deceptive AI in 2024 elections, at the Munich Security Conference. And so, since then, I at least haven't heard a ton about it, I'm sure maybe others haven't either. So I'm curious. Just a few months down the road, if you can share anything, any updates about the status of some of those commitments?

ALTMAN: I don't have the document in front of me, but the things that I can remember. You know, we've already talked about a lot of the elements there. And I do think all of the companies are super focused on the elections around the world this year and doing some of the things that, we've covered so far in this discussion and more to to try to make sure that we can measure what's happening and avoid interference. I, I don't. We're going to be, like, paranoid about it until the elections are done. And then beyond that, there'll

be more elections and there'll be a lot of other things in society, too. But the level of seriousness that the AI companies and the platforms are treating this with. I am very happy to see.

WIRTSCHAFTER: That's great to hear. I mean, just generally speaking, what makes this type of collaboration challenging and what makes it work?

ALTMAN: One of the challenges is, everyone is very careful not to run afoul of antitrust law in coordinating with competitors. And so there's. You know, like you got to have, like, lawyers in the room for the conversations, making sure. And you can't kind of have a casual call for reasons I understand. But that that is like one issue that comes up. And the thing that makes it work is I think most of the people at the leading edge of I really care and have strong opinions about where this is going to go and feel a deep desire to get it right.

WIRTSCHAFTER: So of course, as you mentioned, the elections are one sort of sigh of relief. But this tools, these technologies are here to stay. Information integrity challenges are going to remain difficult well beyond this space. So are there any other opportunities or any other ways, tech companies can think about this collaboration question? Potentially beyond elections.

ALTMAN: I assume we're going to all figure out how to solve, some of these information issues, although they'll be challenging, no doubt. The thing I'm most worried about right now is just the the sort of the speed and magnitude of the socioeconomic impact change that this may have and what the impacts on that will be. That's obviously not. It's a much, much bigger conversation than just the companies. But I feel like there was more conversation about that maybe a year ago. And then GPT four didn't have this, like huge detectable impact on the economy. And so people kind of like, oh, well, you know, we weren't we were too worried about that. It's not a problem. And I have a fear that we just won't take that one seriously enough going forward. And it's a massive, massive issue.

WIRTSCHAFTER: And that's something that I would work across companies as well, considering the the range of competitors of the range of chat bots, the range of options in this space.

ALTMAN: Not only to not just be across companies, but across companies, governments, that that that's like a very broad range thing.

WIRTSCHAFTER: I know we have tons of colleagues at Brookings who would be very eager to have that conversation, particularly about the workforce disruption side. But considering our focus is geopolitics. I'll stick in my lane. Maybe shift gears a little bit. It's clear, obviously, that the AI race is an area of geopolitical competition. We've seen it and in lots of different conversations and in many different ways. In 2022, of course, the Biden administration put in sweeping export controls designed to restrict China's ability to develop and purchase high end chips. Congress passed the Chips and Science Act, designed to boost U.S. manufacturing and R&D around semiconductors. So I'm curious how U.S. government focus on this industry as a clear strategic priority. How does that shape. Open AI, AI development pipeline or process?

ALTMAN: There's a tricky balance to get right here. There's this sort of. The geopolitical realities. And we're like very much clearly on the side of the US and our allies there. And then there's this sort of humanitarian angle of, we want this technology to benefit all of humanity, not not people that not not benefit people who happen to live in a certain country where we don't agree with the leadership and. We don't. We debate this all the time. We don't have like an answer here of how you balance those two things. But, I think they're they are both like. They're both deeply important to get right. And there will be some trade offs.

WIRTSCHAFTER: Do you have a sense already of what those kinds of trade offs might be?

ALTMAN: Where should we be willing to build and operate data centers? Where should we? Not allow services to be accessed for you, even if we would even be able to enforce that at all those sorts of things.

WIRTSCHAFTER: We'll definitely get back. I know to the the data center question, because I know that that's an area of significant interest right now. But kind of sticking with the public private element for a moment. Where do you see potentially the need for additional public investment in AI infrastructure? What do you think about when you say AI infrastructure? Is that truly the data centers? Is it talent broadly etc. But where do you see the need for that public investment to complement or extend address gaps in what the private sector is able to do?

ALTMAN: I think about our infrastructure is mostly compute chips and data centers. There's other things too, you know, new network technologies, whatever else. But I usually mean it is separate from the like AI research talent. I think that. Access to AI compute. So this whole stack of AI infrastructure is probably the

most important commodity in the future, and the best way to make it cheap and thus super broadly accessible. It's to build lots of it. I hope that the private industry can mostly fund it, but I also think it's not a bad thing for the government to fund it as a public good. No matter who funds that, I hope it's like equitably distributed to people to use it. And again, I hope there's like, I hope there's so much of it that we really finally get to talk about intelligence to cheap to meter. The government will need to do some things, even if not write the checks for it. The way we think about permitting, the speed with which we can build this out, the pieces of supply chain that we need to help enable there. There will be an important role for the government there. And, you know, I and many others would really, really like to see this all happen in the US.

WIRTSCHAFTER: So you've previously and I think you hit on this a little bit before, but I think described compute is the currency of the future. Earlier this year, of course, OpenAI drew quite a bit of attention, over efforts to boost global chip manufacturing or ongoing conversations. Compute, of course, has also been highlighted as a tangible asset in a potentially fruitful area for governance. So, yes, and I think you maybe hit on this a little bit, but what is really kind of the current need to increase compute capacity?

ALTMAN: I'm feeling pretty good about effort and progress towards making more chips. I don't think we're like out of the woods yet, but reasonably good. The energy bottleneck that will need to power those chips seems super hard right now. The ability to just build like construct, wire up, you know, rack up these data centers. Also seems hard right now. So, you know, I think we maybe knocked on one challenge or partially knocked out one challenge, and there's no ones in front of us.

WIRTSCHAFTER: So how so seeming in this, in this world where, you know, thinking about lowering obstacles to build data centers in other countries or things like that. How can we address the growing need of AI infrastructure, without questions or compromising potentially questions around privacy or IP or copyright and the like? And then I think that all feeds also into the sort of geopolitical importance of this space, and how that may shape challenges around compute as well.

ALTMAN: Well, I think questions about infrastructure. And you know what we all decide the right copyright framing should be are fairly separate. But the the question of infrastructure and where we build it and you know, who gets to use it and how we think about governance over that. That is a very live and hotly debated question. And like very much above my pay grade.

WIRTSCHAFTER: But I do think that you know, what your what your perspective on that is certainly important, given, I think, how influential OpenAI has been in a lot of these conversations. And given that, you know, you said AI, computing power and and there were thousands of op eds about it or thousands of articles written.

ALTMAN: I mean, I think it's important to like. What I would like. And again, this is like my Sam's perspective. This is not even opening eyes perspective. And this is not a sophisticated expert from this part of the world. But what I would like is for this to be a, US led coalition, but a fairly broadly inclusive one. I don't I don't think it's going to work to say like, oh, only the US should get to build AI data centers. Like, if I were, I don't think the rest of the world would be very happy about that. And I wouldn't feel very good about it either. But I do think finding a way to make sure that we get to, we get some influence over that, and get to help shape that seems important.

WIRTSCHAFTER: That's also, I think, part of, standards question. And I know you've talked a little bit about kind of the global governance side and on standard setting and on regulatory frameworks. So I'm wondering, do you have any thoughts about how, some of these conversations might fit into a broader governance conversation?

ALTMAN: I mean, I can certainly imagine. Some sort of world where there's like a group of people and they say, okay, we, we this technology has global impact. We'll all build data centers. We'll all get sort of. They've heard these like different phrases for it, but like sovereign AI or whatever. But like we've done with. Synthetic biology or nuclear material. There needs to be. Some like rules of the club for. The most catastrophic potential harms. And that's kind of that's kind of the deal in both directions. I could again, deeply out of my area of expertise here, but I could imagine something like that.

WIRTSCHAFTER: So one of the the kinds of rules of the club or maybe an area that I find interesting, I'm not quite 100% sold because I think risk and compute are linked often, but not always. But one of those areas is thinking about regulation around, AI inputs. Or the compute aspect. Something that's maybe more tangible than data or algorithms. From a regulatory perspective. So I think given concerns about some of the pace of AI development, what do you view as sort of the right balance? Well, one, do you think that's a feasible proposition? And two, what do you think is maybe the right balance between boosting that compute infrastructure that you're talking about and questions around regulating these tangible inputs?

ALTMAN: Yeah. I mean, there have been these, like, efforts to say, you know, above this many. Training flops or this much, you know, computing power that there's some other regulatory regime. And I and I appreciate the spirit. I think what the spirit of that is, which is saying, like, we don't want to regulate small startups or open source. We want, you know, we don't want to exclude innovators here. But the most powerful. The most powerful models will be the largest ones. And those can, you know, take more and take more like regulatory over overhead. And I agree with the spirit of that. But I think progress is likely to surprise us in many ways, and you can imagine much smaller models in the future that can still do quite impactful things. It would seem better to me to get to a regime where it's just like, okay, we have these like safety tests. Only they'll only apply to like, quite powerful models. But if you have a model that is capable of you know, harms that these tests would suggest. Then there's going to be some sort of regime for it. And I think, I hope whatever that those things will be. These models will be expensive enough to make and complex enough to make for a while that the people who are doing them will be like responsible, serious actors. So I'd love to get to a world where we have like, sort of this test based framework.

WIRTSCHAFTER: That's great. Yeah. I mean, I think that we have proxies and measures of risk and they can be somewhat accurate or fairly accurate, but they miss things like, oh, of course, there's the the malware writing adaptations and things like that that, that are already able to inflict harm and some of these cybersecurity concerns. Of course, one other area, and maybe after this one question, I will turn it over to my colleague Mike for a little bit. But curious again, because I know in the EO there were questions around cybersecurity, in the voluntary commitments, there were commitments around cybersecurity as well. So and particularly given, the geopolitical interest in this area. Of course, the, the IP elements of it, how does OpenAI think about these cybersecurity breaches and, and what are what are some of the adaptations that you all have been thinking about in this space?

ALTMAN: You know, you. I can definitely imagine a world where we figure out how to, like, align an AI and make it technically safe, and we figure out how to, like, navigate the socioeconomic impacts. And then we have a cybersecurity breach that negates a lot of that. So, we take it extremely seriously. It's a very, very hard problem. I wish I could tell you, like I know the answer, to keep everything we ever do from being stolen by our adversaries. Obviously, I don't know that, but it's something that I spent a lot of my personal attention on.

WIRTSCHAFTER: Yeah, I mean, it would be really nice if there were answers to every aspect in this space, but, you know, it's hard even to predict toward the end of the year. But maybe for now, I'll turn it over for a bit to my colleague Mike O'Hanlon, who I know is quite eager to ask you some questions about the recent application of AI systems beyond generative AI, to areas that are kind of ongoing national security questions, and concerns including in both Gaza and Ukraine. So we're also going to be monitoring the chat or the email or Twitter or wherever. You're submitting your questions. We're going to be monitoring that for incoming questions from the online audience as well. So please do submit them. And I think, and I hope we should have some time to get to those before the end of the session.

O'HANLON: Thanks, Valerie, and thank you, Sam and Valerie for a great conversation so far. By the way, I think the email address is events@brookings.edu if folks still want to send additional questions. But if I could let me begin before I get into a couple of questions on war and peace and defense technology with making sure that I'm integrating some of the previous conversation correctly. So I just wanted to throw this back at you. You mentioned that you're kind of certainly be vigilant through November and beyond in the ways in which, artificial intelligence could be used and misused and also productively used, to make these elections in the United States and elsewhere as safe and fair as possible. But it strikes me that so far, you seem hopeful that we can do a pretty good job. And the thought that occurred to me is that maybe the 2024 elections won't be as bothered by cyber problems writ large as the 2016 elections and I want to see if you agree with that or if you even want to comment, because in 2016, we had such a close election between Donald Trump and Hillary Clinton that even small vote swings could make a huge difference. That may still be true in the upcoming election, but we were, I think, a little bit as a when I say we the broader country, we're caught a little bit unawares about the ways in which social media could be misused through trolls and foreign agents that we hadn't really thought a lot about before in the general conversation. And so there wound up being perhaps a greater impact, as people weren't quite as suspicious about what might be done by, let's say, Russian trolls. And on top of that, of course, we had a big email leak, which, revealed a lot of inner conversation and was therefore, you know, hurtful probably to Hillary Clinton's campaign. Whereas what's new about AI in this election cycle, if I'm hearing you right, is the ability to tailor messaging to individuals, perhaps, and that could still be very consequential. But it's going to happen in person by person and small group by small group. And therefore, perhaps we can be hopeful that cyber problems broadly defined to include AI, but not exclusively, maybe less deleterious, less harmful to this upcoming election than they were eight years ago. Is that a fair overall conclusion? I know it's too soon to say for sure.

ALTMAN: Not not not really. I mean, I don't want to I don't want to say like, I don't -- what I was trying to say earlier is I think people have a higher level of vigilance. I don't I think people won't make the same mistakes as last time. But there may be new threats there. There may be unknown unknowns that we don't. Realized until too late. I also personalized one on one. Persuasion is actually scarier to me. Not less scary than the mask stuff. It's harder to detect it and probably more powerful. And again, and I can talk to a lot of people at once. So I don't want to give I very much don't want to give a sense that we're not worried, but I think it's good that we are worried. And that's leading to that's leading to vigilance. In the industry.

O'HANLON: Right. So I wanted to talk a little bit about defense technology, if I could, and not with a specific focus on anything you may be doing there or any other company, but just generally about trends in military technology. I know you've been concerned about AI and the possibility of nuclear war, and trying to make sure that there's a human in the loop in any future decision on the use of nuclear weapons. We see recent reports that China and the United States may wind up talking about this very question officially. And so if I imagine the future applications of AI within weapons systems, on the one extreme, it sounds like there's a movement towards consensus that artificial intelligence should never decide to initiate a nuclear attack, and that maybe we can work towards a world in which that norm is agreed to and backed up with various kinds of technical safeguards. On the other extreme, however, I could imagine certain kinds of swarms of robotics used in a defense where you absolutely would want to have AI deciding when to shoot because humans can't do it fast enough. So let's say, hypothetically, a bunch of quadcopters in Seoul, South Korea, in the face of a North Korean artillery and missile barrage. There's nothing wrong that I can see with having some future swarm of robotic defensive vehicles shoot down those incoming North Korean rockets and artillery shells and reposition themselves to maintain a robust net, using artificial intelligence or other advanced algorithms. Just because they need to be fast enough to deal with this barrage. So to me, the hard part is in between. And I wanted to ask you if you agree. But as one hypothetical in-between, let's imagine North Korea now launches 100 airplanes coming south towards Seoul. They probably wouldn't because they're not known for their air power these days. But let's say we had a defensive swarm of robotic vehicles powered and interconnected through an AI system that could be empowered to shoot down those manned aircraft and kill 100 North Korean pilots. It's a defensive operation, but it's leading to lethal, you know, activity and loss of human life. To me, that's sort of a case where maybe you would use AI, but it gets a little bit dicier. So do you agree with the way I framed this and where are you along that spectrum?

ALTMAN: Yeah, as you said, I've never heard anyone advocate that AI should get to make decisions about launching nuclear weapons. I've also never heard anyone advocate that I shouldn't be used to, like, intercept. You know, inbound missiles where you have to act really quickly. And then there's this whole area in the middle of, well, if it's, you know, if there's like a plane coming to bomb South Korea and you don't have time to have a human in the loop, and you can make an intercept decision or not, but like you're very sure that it's happening, like how short you have to be. What would be the expected impact of human life? And you know, where on that? Like where do you draw the line in that gray area? And man, there's like a lot of questions. I think I could provide you with some expertise on if I. I don't think this should be. I hope this is never an OpenAI decision. If somehow it were like you'd be one of the people I'd call about it, I'd call a bunch of, like, experts from the defense industry who have thought about this their entire career, but I don't think I have, like, much expertise or nuance to really sort of weigh in an interesting way on that one.

O'HANLON: Well, I'm glad at least, that my framework doesn't strike you as crazy for how to think about the question. And as you say, it's going to be the devil in the details about all the questions. Like you say, confidence level, what's the potential loss of life if you're wrong, etc.. By the way, I wanted to compliment you because, I, I use ChatGPT 3.5 to, pose my final exam questions. I'm teaching two courses at Columbia right now, and I hope my students aren't listening because, ChatGPT 3.5 gave pretty good answers. That would probably be

ALTMAN: Have you tried it with four?

O'HANLON: Well, I haven't yet tried it with four. So, I probably should because some of my students probably will.

ALTMAN: Probably they do.

O'HANLON: Well as well. By the way, I'm sure you know that universities are wrestling with, how much to let people use ChatGPT. And my answer. But I want to hear yours, if you don't mind, before I come back to another question on war and peace, my answer is as long as you footnote it, you know it's probably the single best reference in the world. Why would you not want to let people footnote the encyclopedia? But ChatGPT is better than the encyclopedia. It's more up to date. It's more nuanced. So I could never prevent a student from using any written reference material, so I'm not going to try to prevent them from using

ChatGPT, but they got to tell me what they're doing, and if they want, they can ask ChatGPT of any vintage to answer the question and then critique the answer that ChatGPT just gave, and do that as an alternative final exam. So what do you think about this?

ALTMAN: Yeah, I mean, I understand, I agree, I not only I though I think you should allow well, I have a bias here, but not only that, you should allow students to use it. I think you should require it. I think you know. Education needs to prepare people for the world they're going to inhabit and hopefully be productive in, as adults. And humans have always built better and better tools that new generations use to go on to greater and greater heights. There were obviously times where teachers tried to ban calculators or as you mentioned, Google search. And that I think it's not a winning strategy. People in the future are for sure going to be using AI to help them do better work. And they may as well learn how to use them. So I suppose that that's all great, but but but use the tool.

O'HANLON: My big picture question I wanted to ask you about war and peace is whether you think that I will likely make war more or less likely in the future, and more or less severe and deadly if and when it happens. And if I could just frame the way I think about it. I mean, for me, it's going to be we don't know yet because it's going to depend on partly how we come to understand AI strengths and potential liabilities as a human species, but also on the way individual people will use it. I could imagine, for example, some future leader contemplating the possibility of aggression against a neighbor. And he's this leader is armed with the latest ChatGPT, or maybe even thinks that he's hired a scientist that's done better than the latest ChatGPT, and he thinks he can come up with a war plan that his would be adversary would not have yet figured out. And therefore, based on this false allure of technology, which, by the way, is often historically what leads to war, people think that they could use a new technology better than their adversary, and that leads them to think they have a war plan that will work quickly and lethally and well and successfully. And so they wind up getting overconfident, and they're usually wrong. And you could imagine the same thing happening with any kind of cyber technology and certainly AI. But you could also imagine that this leader will worry that his potential victim has pretty good AI himself. And even if the potential victim is not that smart in the mind of would be aggressor country A, that he's now going to be able to anticipate the various types of ambushes or attacks that could befall him. And therefore this sort of neutralizes the the incentive to attack. It puts people on a more level playing field and reduces the likelihood of an attack, because, again, there's less of a probability of a successful surprise or successful, you know, new innovative concept. Am I just speaking gibberish? It does this do you think like this at all? And where do you come down?

ALTMAN: Again. We're pretty far from my area of expertise here. I, I mean, I think there's like a bunch of different factors and I don't know how to wait, how they're all they're all going to play out. I, I do think generally as the world gets more prosperous, more interconnected, and particularly as the like rate of increase, annual increase in prosperity goes up, that it seems to me historically, you know, not perfectly, but as a general correlation that. We have less war. And, I hope that if I can be a huge, like, boon to prosperity and abundance. That will be one impact. Then there's the things that you said, which I don't even know how to, like, net out what those are, what the impact of that is going to be. And then there's this other thing, which is if this feels like it's going to reshape the geopolitical landscape of the world. That's like a fairly high stakes. And I don't know how different countries respond to that.

O'HANLON: That's a really good framework. So yeah, think about trends towards prosperity and also globalization, which has not always been a panacea for preventing war. But like you say, the overall trend line has been at least generally favorable over the last hundred years. And then there's the question of sort of wargaming out who's going to think, what plan might work. And then there's, like you say, shaping geopolitics. So thank you. That's a good framework. That's a helpful framework for me to think about as well. I'm now going to toss back to Valerie, who's got, I'm sure I think we had at last count into the hundreds of audience questions lined up. And so she's going to give some, select version of, of those and maybe add 1 or 2 more of a round to round out the hour. So. Thank you.

WIRTSCHAFTER: Yeah. So, I'm sorry. These might be a little bit scattered just because they're coming in from different places. So I've tried to, to bunch I've been typing in, bunching questions furiously. But, it's a little difficult while trying to listen to. So, I guess first set of questions, maybe thinking about this open source question and the implications of that for geopolitics. So originally, of course, OpenAI was largely transparent about algorithms, training, data processes. What what sort of behind the, the the closing or the gradual closing down? And then, second question to later on today, that's more of my own is I saw yesterday, actually, that there was some progress from Chinese researchers, and thinking about a way to openly release models that are hard to fine tune. And so I'm wondering how that potential evolution could reshape if it's possible or or not might reshape, the calculus in terms of open models or open weights or, or some form of openness on the spectrum of, of what it means to be open source.

ALTMAN: Yeah. Look, I think there's a place for open source models or open weight models. I think there's a place for large proprietary models, in the cloud. And I support both. I think we should we will, we do open source some stuff. We'll continue to open source more. I hadn't seen that work about the model. That's hard to find soon, but I'm very curious, so I'll try to find it after this.

WIRTSCHAFTER: I'll send it to you. I mean, yeah, it's it's an interesting, possibility. I, I haven't looked deeply at it, but, it it seems like if it, if it's harder to make these sort of malicious, adaptations that could be promising. Of course, but, and might reshape the, the evolution in terms of thinking about open sourcing. So I don't know if even if you have a thought about that or if you're going to wait and share your thought at some point when you've had a chance to, to take a look. But.

ALTMAN: Yeah, I don't think I can react too much till I take a look, but it's it's it sounds interesting.

WIRTSCHAFTER: Yeah, it was definitely quite interesting. So I guess in another bunch of questions, my transitions are flawless here. But thinking about the China question, of course. That's always a challenge in terms of global governance conversations. Kind of thinking about the, the ways to get groups of countries together. So what role do you see China playing in a lot of these conversations? What area might be the easiest to find consensus and where do you think it might be most difficult? Particularly around and some effective regime on global governance enforcement for AI.

ALTMAN: I mean, I hate to keep saying you're asking the non-expert in the room, but I feel like so embarrassed to try to answer this question with this audience. Like, I think it's a super important question. The kind of people that I've been calling to ask for input on it are much more like the people on this call than me. I am. I am generally hopeful that although we will disagree with many, many important things. With the Chinese government about AI that we all do share a, goal of reducing catastrophic risk. And I think we should try really hard to see if we can come to an agreement on some of the highest level safety issues there.

WIRTSCHAFTER: I think. Yeah. I mean, I wholly agree. I mean, I think that some of the, the conversations that have been ongoing with respect to China around AI, usage have been focused on the sort of one definitionally, which has always been a challenge. And I'm curious, your take on just broadly the difficulties of getting definitions and norms or expectations of what things actually are, at an international level, whether

you found that to be difficult. But I think definitionally there have been questions around shared, sort of. What are we talking about when we're talking about AI? But then also where the the minimum is. And I think a lot of that has been around conflict, usage and especially that catastrophic risk component. But yeah, I don't I don't know if you have any thoughts about sort of just the challenge of just definitionally, coming to consensus at the international level.

ALTMAN: I know there's like many. There's many form dialog going on about how we come to those definitions and where we want to focus. I don't know what the outcome is of the current status of them is.

WIRTSCHAFTER: So I guess thinking then on the governance side still, but maybe a bit more domestically. We saw a lot of momentum, around, whether it was the I insight forums, which I believe you participated in. There were tons of bills proposed. You know, I'm a bit and I know one of the things that you have been sort of vocal about is we need to think about this, and come to some sort of. Regulatory regime of some kind. I don't know what your perspective is on the current state of conversations. If you have any, sort of, sort of progress to report in that area.

ALTMAN: I don't have anything that is not that's like nonpublic there. I think people keep talking, and it's not clear to me what the path to getting something enacted at this point looks like. I think there's many important things to do, just sort of given our mission and our focus on AGI. We are particularly interested in the questions of catastrophic risk, so that that's what I've been tracking the most personally, even though I think there's like tons of other important parts of this too.

WIRTSCHAFTER: Do you think that so in terms of the kind of balance of risks versus the sort of AGI component and the here and now aspects of risks? What do you think is the right balance to think about in terms of regulatory, elements of that conversation?

ALTMAN: I mean, you've got to address both. I think it would be amiss. Like. You kind of. You do sort of get two camps there and many people are like quite dismissive of the other. And I think that's a real mistake. Like we've got to get it right at every stage.

WIRTSCHAFTER: And I know one of the sort of here now risks that we'd heard quite a bit about, was, of course, the election context and deepfakes of, can candidates. And like, I know OpenAI has, rules around

campaign usages of AI. I was curious, you know, maybe not in a US context, but in other contexts. I can be sort of that. I think you've described it as the brainstorming partner. I view it as, sort of the. If there's an inequality to be able to bridge some of that gap in terms of resources potentially. So I'm curious about just the usage of open AI tools or generative AI broadly. On the positive side of elections, if you have any thoughts about that and how that maybe comes into contrast with some of OpenAI's policies.

ALTMAN: Yeah, I think there are clearly positives. With a principle of conservatism and what we know or caution and what we know right now, I think it is easier for us to see. The importance of. Being more cautious about mitigating the downsides, then more permissive in allowing the upsides for like political campaign use. Although, of course there are upsides too. And, you know, this is a judgment call and this is where we head out right now.

WIRTSCHAFTER: So a few more questions coming in. They're big ones. Just long. So I'm trying to read through to get to the question, but, I think going back to, to Sora and we talked a little bit about some of the, the thinking behind, Sora. I think OpenAI, is planning to publicly release it this year. Is that correct?

ALTMAN: I don't think we've committed to an exact timeline. But.

WIRTSCHAFTER: So one of the questions was, why not hold off until at least after this election season passes? And then also, what kinds of the watermarks and other safeguards are going to be built in? And maybe I'll throw in the The voice tool as well to Sora. So thinking about the, the technical solutions for both Sora and the, the voice, tool that you all have built as well, what are some of those things that are going to be thrown in there?

ALTMAN: Yeah. So I don't think we've, said any intention about releasing something before an election this year. I do think other people may and so the world should be prepared for these tools to be out there. But I don't think we've made any statement about an intention for that. You know, we're working on lots of watermarking things, and we've talked about some of that work. I do want to flag something else that I think is underexplored, which is the idea not just of watermarking generated content, but authenticating non generated content. And so this idea that if someone in an election has a really important thing message to say, they cryptographically sign it and say hey I said this and here's the proof and you can check that. That

seems to me like a reasonably likely part of the future, for certain kinds of messages. And I think we should talk more about that.

WIRTSCHAFTER: Yeah, I think that was part of the the EO as well, was to have official statements, have some sort of cryptographic signature. Mike, did you have something you wanted to?

O'HANLON: Yeah, I was gonna try to. Sorry to interrupt. I was going to try to throw in one more big picture question. Give you a chance, Valerie, to maybe read through the last 1 or 2 that you want to post.

WIRTSCHAFTER: They're really, as we really long everybody.

O'HANLON: But but, Sam, I know how, how hopeful and bullish you are on AI with all the caveats that we've been talking about today, all the things you're working on. And I guess I wondered if you wanted to give a message to the typical voter about how to think about AI in this election cycle, because we've been talking about high level policy primarily in this conversation. But I wonder how an individual voter should view the not just the perils, but also the opportunities of AI as people try to make sense of who to vote for. This year, I realize a lot of it's going to be a personality Rorschach test with Trump and Biden. So it's a little different election in some years. But obviously the issues do matter and ideas matter, and I can help. Right. So do you have any message you would offer to a voter? Leave aside the big policy questions and debates.

ALTMAN: Well. I do think I should be a top of mind issue for every voter in this election. The impact it's going to have on the economy and how we want to position that, and also making sure that we have U.S. leadership and that the US is a major winner of this new revolution. I expect that that will be an issue in this election. Maybe not as much as it should be, but I would very much encourage people to be quite focused on that.

WIRTSCHAFTER: I think that's why these types of conversations are also really great. So thank you for for joining us. So I, I went back into the, the long questions. So we have maybe time for a couple more. So we'll shift gears a little bit back to where we were before and thinking about the electricity demands. And I know that you've, talked a bit about electricity demands from data centers. So it's also it's unclear, what some of these advances will look like. And so thinking about the electricity side of things, how can OpenAI and other tech companies involved in this space think about or square investments in expanding these AI services, and

data centers with, some kind of responsibility, to help reduce carbon emissions, mitigate the impact of climate change. Thinking about those bigger global issues, of course, that impact lots of countries and people around the world. At a time when, we've been experiencing, of course, those disruptive climate related impacts.

ALTMAN: One of the things that I'm most excited about for with I like many wonderful things, you know, great impact on education. Awesome. But the fact that I can help us do make more scientific progress, come up with new understandings and ways to combat climate change, faster transition to clean energy, think about new ways to do carbon capture. That's great. And more broadly, the idea of using AI tools for faster scientific progress is probably the area that is personally, to me, most exciting.

WIRTSCHAFTER: So maybe we have time for one more, two more questions. So, I'll I'll throw one out there. That's in Mike's domain. So I apologize, Mike, for, you know, if I, if I misstep. But I guess maybe you touched a little bit of on upon this, but thinking about the positives, how can we think about AI maybe making the world more peaceful? Or maybe making diplomacy more possible?

ALTMAN: More prosperity and better understanding of each other and identify new win win solutions. I believe all of that should work.

WIRTSCHAFTER: Okay, well, with that answer, we have time for, I think, more questions. So, Mike, I don't know if you had any others that came to mind for you, but I'll I'll poke around here and see if there any others that. Sure.

O'HANLON: I'll throw my last one in while I add my thanks because I know Val, you're going to you're going to sign off with Sam with your last question. So, Sam, thank you so much. I guess my other question related to the election is, has ChatGPT gotten good enough where a curious voter who wants to understand issues better can just routinely ask ChatGPT 4.0 for how to understand the options in improving American health care or improving, you know, gun safety or reducing crime or improving immigration policy? Are we at a point where that's the go to place to just ask, or should voters wait another four years for that?

ALTMAN: I wouldn't say it's the go to place, but I think it can provide some assistance there.

WIRTSCHAFTER: Kind of I think building on that, we're seeing, of course, chat bots get built into search engines. We know there's a lot of research. Of course, there's not a lot of research on on many of the things we've been discussing, yet. And I hope there will be a lot more, especially around the persuasion elements that you've highlighted is particularly concerning. But we've seen, of course, chat bots get built into search engines. We know people don't go much past the first, page. We know that they don't necessarily read past the headlines. So I'm curious, if, if this sort of deployment are ready for prime time, challenges that you've highlighted, what that sort of means for the information space broadly to have these start to become part of that place where we do our own research.

ALTMAN: I mean, I don't think we know the answer to that yet. There will. For sure I expect these tools to change how people find and synthesize and use information, and we're seeing some of that. Already, but the and this is maybe a nice thing to close with. We're still so early in the arc of this technology and the co-evolution of society and this technology that I don't think we can say like here. So everyone's going to. Do their work and find their information in another few years. Like, we'll have to just have a very tight feedback loop in. Watch what people want to do. And try to support that.

WIRTSCHAFTER: I think that's great, and I feel like that fits into the way that OpenAI has approached sort of incremental adaptation of its tools, which are, of course, extremely impressive. I'm heartened to hear, really the the sort of thought and, pause that has gone in, especially into some of the, the tools that can, I think, be more misused than others, like voice cloning. So I guess, you know, I will, unless you had any other pressing thoughts that you wanted to share with our audience today. I will end my line of questioning. And thank you very, very much for taking this time to to talk to us. It it's wonderful to to meet you and to to hear your thinking in this space. And I hope we'll be able to to do it again. As these tools continue to evolve, maybe it'll have to be next month, given how fast they are evolving. But, I certainly will be be following along and, and and very grateful. For the, the time that you've given us today.

ALTMAN: Thank you very much.

O'HANLON: Thank you. Sam.