Center for Humane Technology | *Your Undivided Attention* Podcast

Feed Drop: AI Doomsday with Kara Swisher

Tristan Harris: Hey everyone, it's Tristan. Recently I sat down with veteran tech journalist Kara Swisher for her podcast *On with Kara Swisher*. And we thought about sharing that interview directly here with you on *Your Undivided Attention* because, as you'll hear, there's really no one better than Kara to challenge people to articulate what's really going on about a situation. A lot of people have called Aza and I fearmongers or doomsayers. And the point of our AI Dilemma presentation is not to sow fear or doom, it's to say we have to honestly assess the risks so that we can choose to take the actions that are needed to avoid those risks. And I think this interview did a really great job of distilling a lot of our current thinking on AI since the space is moving incredibly fast.

So if you're a new listener or you want to send this to friends, family, or broader network, it's a great way into the AI topic. And if you like it and you want to hear more of Kara's interviews with folks like Sam Altman, Reid Hoffman and others, go to wherever you're listening to this podcast and search for *On with Kara Swisher*. And now over to Kara.

Kara Swisher: Welcome, Tristan. Now you and I met, let's go back a little bit, when you were concerned about social media. I think it was one of the first interviews we did.

Tristan Harris: It was 2016, 2017.

Kara Swisher: Right.

Tristan Harris: I think it was right after Trump had gotten elected.

Kara Swisher: That's correct.

Tristan Harris: And I was really choosing to come out and say...

Kara Swisher: In a little booth in Stanford, I remember.

Tristan Harris: I remember that. The Stanford radio.

Kara Swisher: It was very small, yeah.

Tristan Harris: Yeah. Mm-hmm.

Kara Swisher: Talk about for people who don't know you, both of us are probably seen as irritants or Cassandras, I guess, she was right.

Tristan Harris: Yeah.

Kara Swisher: Whatever, John the Baptist; any of those precursors lost his head, okay. Talk about what got you concerned in the first place, just very briefly for people to understand.

Tristan Harris: So I guess for people who don't know my background, I was a tech entrepreneur. I had a tiny company called Apture. We got talent acquired by Google. In college I was part of a class called the Stanford Persuasive Technology Lab Behavior Design Class. And studying the field of social psychology, persuasion and technology, how does technology persuade people's attitudes, beliefs, and behaviors.

Kara Swisher: Right.

Tristan Harris: And then I saw how those techniques were the mechanisms of the arms race to engage people with attention. Because how do I get your attention? I'm better at pulling on a string in the human brain, in the human mind. And so I became a design ethicist at Google after releasing a presentation inside the company in 2013, saying that we were stewarding the collective consciousness of humanity. We were rewiring the flows of attention, the flows of information, the flows of relationships. I sort of said, "I'm really worried about this." I actually thought the presentation was going to get me fired. And instead I became a design ethicist going to study, how would we?

Kara Swisher: They gave you a job of these worries, right?

Tristan Harris: Yes. It's better than me leaving and doing something else.

Kara Swisher: Right.

Tristan Harris: But I tried to change Google from the inside for three years before leaving.

Kara Swisher: When you look back on that, I think they wanted to have you there. You're kind of like a house pet, right? You know what I mean? "Oh, we got to design that."

Tristan Harris: Hopefully I'm more friendly than a house pet.

Kara Swisher: Yeah, I know. But they don't like the house pets that bite. And you started to bite.

Tristan Harris: Yeah. Well, it's funny now because if you look at, when we get to AI, which we're going to get to later, people who started AI companies actually started with the notion of, "We can do tremendous damage with what we've created." There's a whole field of AI safety and AI risk.

Kara Swisher: Yes.

Tristan Harris: Now, imagine if when we created social media companies, Mark Zuckerberg and Jack Dorsey and all these guys said, "We can wreck society. We need to have a whole field of social media safety, social media risk." And they had actually had safety teams from the very beginning figuring out...

Kara Swisher: They hated when you brought up negative things.

Tristan Harris: They denied that there was even any issue. And it was hard to see the issue. And we had to fight for the idea, you and I, that there was these major issues. Addiction, polarization, narcissism, validation seeking sexualization of kids, online harassment, bullying.

Kara Swisher: Yes.

Tristan Harris: These are all digital fallout of the race to the bottom of the brainstem for attention, the race to be more and more aggressive about attention.

Kara Swisher: Right.

Tristan Harris: So I was frustrated that, especially Facebook, because I had more contact with that company, wasn't going to do more. And that people were in denial about it. And it goes back to the Upton Sinclair quote, "You can't get someone to question something that their salary depends on them not seeing."

Kara Swisher: That's true. Yeah. And their boss, their boss who runs everything. It's like a brick wall on that.

Tristan Harris: Yeah. We're a neutral mirror for society. We're just showing you the unfortunate facts about how your society already feels and works.

Kara Swisher: Yeah. I kept saying finish college, you'll understand.

Tristan Harris: Yeah.

Kara Swisher: You might want to take World War II maybe.

Tristan Harris: Yeah.

Kara Swisher: Throw in some Vietnam War and perhaps go back to World War I, because that's recent history.

Tristan Harris: Yeah.

Kara Swisher: So a couple of months ago, you and Aza had a presentation that I went to here in Washington called The AI Dilemma, laying out the fears. I think there's a proclivity to say, "Calm down, don't be so terminator." There's a proclivity to say, "Don't be so sunshine. Let's focus, not on the existential fears, but the ones we can work on." Now, one of the people that have been working on it, feel like you can't guess what it's going to do at this point. And that when you get overly dramatic, it's a real problem. Yours was pretty dramatic when you were doing it in front of a group of Washington people.

Tristan Harris: When you say you can't guess what, what do you mean?

Kara Swisher: What's going to happen with this? We don't know, so let's deal with our current fears versus our supposed fears.

Tristan Harris: Yeah, I disagree. So first of all, there's a whole bunch of harms of AI and all the stuff around bias and fairness and automating job applications and police algorithms and loans. And those issues are super important. And they affect the safety of society as it exists. I think the things that we're worried about are the ways that the deployment of AI can undermine the container of society working at all. Cyber attacks that can break critical infrastructure, water systems, nuclear systems, the ability to undermine democracy at scale.

Kara Swisher: In Silicon Valley, it's common for AI researchers to ask each other what their, it's called P(doom), the probability of doom.

Tristan Harris: Yeah. Yeah.

Kara Swisher: Explain what P(doom) is calculating, and tell me what's your P(doom).

Tristan Harris: So I don't know if I have a P(doom). And you were sort of, I want to make sure I go back to the thing you were saying earlier, can we predict what's going to happen? I would say we can predict what's going to happen. And I don't mean that it's doom. What I mean is that a race dynamic where if I don't deploy my AI system as fast as the other guy, I'm going to lose to the guy that is deploying super fast. So if Google, for example...

Kara Swisher: That's internally capitalist companies, and then also other countries.

Tristan Harris: Yes, exactly. And that's just a multipolar trap, a classic race to the cliff.

Kara Swisher: Right.

Tristan Harris: And so Google, for example, had been holding back many advanced AI capabilities in the lab, not deploying them because they thought they were not safe.

Kara Swisher: Yeah.

Tristan Harris: When Microsoft and Open AI hit the starting gun and said in November, "We're going to launch ChatGPT and then boom, we're going to integrate that into Bing. We're going to make Google dance," as Satya Nadella said.

Kara Swisher: Right.

Tristan Harris: That hit the starting gun on a market, a pace of market competition.

Kara Swisher: Right. They have to.

Tristan Harris: Then now everybody is going, "We have to." And we have to what? We have to unsafely, recklessly deploy this as fast as possible.

Kara Swisher: So that we are out front. Like my Google just asked me to write an email. They usually want to finish sentences now.

Tristan Harris: Right.

Kara Swisher: They're like, "Can I write this email for you?" I was like, "Go fuck yourself. No, I don't want you to."

Tristan Harris: Right.

Kara Swisher: They took it well.

Tristan Harris: And then Slack has to integrate their thing and integrate a chatbot, and then Snapchat integrates My AI bots into the way that it works.

Kara Swisher: Spotify.

Tristan Harris: TikTok; and I haven't even seen the Spotify one. I mean, the point is, this is what I mean by you can predict the future. Because what you can predict is that everyone that can integrate AI in a dominating way to become, in the case for the race to engagement in AI, it's the race to intimacy; who can have the dominant relationship slot in your life. If Snapchat AI has a relationship with a 13-year-old that they have for four years, are they going to switch to TikTok or the next AI when it comes out? No, because they've already built up a relationship with that one.

Kara Swisher: Unless AI is everywhere and then you have lots of relationships, like you do in life.

Tristan Harris: Right. But what they'll want to be incentivized to do is to deepen that relationship, to personalize it, to have known everything about you and to really care about you.

Kara Swisher: You want to leave me now?

Tristan Harris: Don't leave me now. And I mean even Facebook did that when you wanted to delete your account in 2016, they would say, "Do you really want to leave?" And they would literally put up photos of the five friends and they would calculate which five friends could I show you that would most dissuade you from doing that.

Kara Swisher: Yeah, yeah.

Tristan Harris: And so now we're going to see more and more sophisticated versions of those kinds of things.

Kara Swisher: Yeah.

Tristan Harris: But that race to intimacy, that race to become that slot in your life, the race to deploy, the race to therefore move recklessly, those are all predictable factors. So just to be clear, because you're sort of challenging me, can we predict where this is going? And the point is we can predict that it was going to go so recklessly and go so quickly, because we're also deploying this faster than we deployed any other technology in history.

Kara Swisher: That's correct.

Tristan Harris: So the most consequential technology, the most powerful technology we've ever deployed, and we're deploying it faster than any other one in history. So for example, it took Facebook four and a half years to get to a hundred million users. It took TikTok nine months, it took ChatGPT, I believe two months.

Kara Swisher: And they have the app now.

Tristan Harris: Yeah.

Kara Swisher: But in the presentation, in that vein, you cite a study where 50% of AI researchers say that their P(doom) is 10% or higher, but it's based on a non peer-reviewed survey, on a single question survey, they had only about 150 responses. Should we be swayed by that data that they're worried? Because there is that ongoing theory that the people who make this are worried, the cooks are worried about what they're making.

Tristan Harris: Yes. So one critique of that survey is it's somehow all about AI hype; that the people who are answering those surveys are people inside the companies who want to hype the capabilities so that they get more funding and that everybody thinks it's bigger than it actually is.

Kara Swisher: Sure.

Tristan Harris: But the people who answered that survey were machine learning researchers who actually publish papers and conferences. They're the people who actually know this stuff the best.

Kara Swisher: Sure.

Tristan Harris: If you go inside the industry and talk to the people who build this stuff, it's much higher than that survey is. Again, this is why we're doing this.

Kara Swisher: Oh, I was at a dinner party years ago when they were top people. I was like, "Huh. That's interesting."

Tristan Harris: Yeah. They're very top people. I mean, don't trust the survey, trust... There's a document of all the quotes of all the founders of AI companies over all the years of saying these quotes about there's a strong chance we'll wipe out humanity, we'll probably go extinct. They're not talking about jobs, they're talking about a whole bunch of other scenarios.

Kara Swisher: Right.

Tristan Harris: So don't let one survey be the thing. We're just trying to take one day at a time.

Kara Swisher: People are worried.

Tristan Harris: People are deeply worried.

Kara Swisher: You use the metaphor of a Golem. Explain the Golem.

Tristan Harris: So the reason that we actually came up with that phrase to describe it, is that people have often said, and this is pre GPT-4 coming out, "Why are we suddenly so worried about AI? AI has existed for 20 years. We haven't freaked out about it until now. And Siri still mispronounces my name and Google Maps still my pronounces the street address that I live on wrong. And why are we suddenly so worried about AI?"

And so one of the things that in our own work, in trying to figure out how we would explain this to people, was realizing that we needed to tell the part of the story that in 2017, AI changed, because a new type of AI, class of AI came out called transformers. It's 200 lines of code, it's based on deep learning. That technology created this brand new explosive wave of AI that are based on generative, large language, multi-modal models; G-L-L-M. We said, "How can we differentiate this new sort of era of AI that we're in from the past, so that people understand why this curve is so explosive and vertical. And so we said, "Okay, let's give that a name so that people can track it better." As public communicators, Aza and I cared deeply about precise communication. So we just said, "Let's call them Golem-class AI's."

Kara Swisher: And a golem of course is the famous...

Tristan Harris: The Jewish myth of an inanimate object that then gains animate capabilities. And that's one of the other factors about generative large language models, is that as you pump them with more information and more compute and you train on them, they actually gain new capabilities that the engineers themselves didn't program into them.

Kara Swisher: Right, that they're learning.

Tristan Harris: They learn things, right.

Kara Swisher: Now, let me be clear. You do not believe these are sentient.

Tristan Harris: No. And this has nothing to do with what their..

Kara Swisher: Make that clear. They're not humans.

Tristan Harris: There's this fascinating tendency when human beings think about this, where they get obsessed with the question of whether they can think.

Kara Swisher: Sci-fi.

Tristan Harris: Sci-fi.

Kara Swisher: That's why.

Tristan Harris: Yeah. But it actually kind of demonstrates just kind of like predispositions of humans. You so imagine Neanderthals are baking homosapiens in a lab and they become obsessed with the question when it comes out, when this thing is more intelligent, is it going to be sentient like Neanderthals? It's just a bias of how our brains work.

Kara Swisher: Right.

Tristan Harris: When really what really matters is, can you anticipate the capabilities of something that's smarter than you? So imagine you're in Neanderthal, you're living in a neanderthal brain. You can't think about humans once they pop out inventing computation, inventing energy, inventing oil-based hydrocarbon economies, inventing language.

Kara Swisher: Right. So we don't know. What you're essentially saying, we don't know.

Tristan Harris: Yeah.

Kara Swisher: It's inconceivable what it is, but it's not sentient. Because then we attribute emotions to it.

Tristan Harris: Right. Well maybe eventually those questions will matter, but they're just not the questions that matter. Whether or not it is sentient; it doesn't have to be.

Kara Swisher: No.

Tristan Harris: There's enormous dangers that can just emerge from just growing these capabilities and entangling this new alien intelligence with society faster than we actually know what's there.

Kara Swisher: Alien is an interesting word that you use, because it's one that Elon Musk used many years ago. He said, "They treat us like aliens would treat a house cat." But then he changed it to, "We are an anthill and they're making a highway." They're not mad at us.

Tristan Harris: No.

Kara Swisher: They don't care.

Tristan Harris: No. They're just doing things from their perspective. That makes sense.

Kara Swisher: Makes sense.

Tristan Harris: But it's just like, by the way, just like social media was.

Kara Swisher: Right.

Tristan Harris: So social media already, let me argue that AI might have already taken control of humanity in the form of first contact with AI, which is social media. What are all of us running around the world doing every day? What are all of our political fears? What are all of our elections? They're all driven by social media. We've been in the social media AI brain implant for 10 years. We don't need an Elon

Musk brain implant. We already have one. It's called social media. It's been feeding us the worldviews and the umwelts that define how we see reality for 10 years.

Kara Swisher: And the noisiest people.

Tristan Harris: And the noisiest people. And that has warped our collective consciousness. And so are you free if all the information you've ever been looking at has already been determined by an AI for the last 10 years? And you're running confirmation bias on a stack of stuff that has been pre-selected from the outrage selection feed of Twitter and the rest of it. And so you could argue that AI has already taken over society in a subtle way. I don't mean taken over in the sense that it's values are driving us, but in the sense that just like we don't have regular chickens anymore, we have the kind of chickens that have been domesticated for their meats. We don't have regular cows, we have the kind of cows that have been domesticated for their milk and their meat. We don't have regular humans anymore. We have AI engagement optimized humans.

Kara Swisher: So one of the things you and Aza did was you made a lot of news when you tested Snapchat's AI, it's My AI, as if you were a 13-year-old. It gave him advice how to set the mood for sex with a 35-year-old. Stunty. They've fixed it. They think they've fixed it.

Tristan Harris: Aza tested it a few days ago; it still happens.

Kara Swisher: It still happens.

Tristan Harris: It is suggesting you bring candles for your first romantic time with a 13-year-old with a 38 or 41-year-old, I think it was.

Kara Swisher: Right.

Tristan Harris: So it doesn't say a couple of the suggestions, but it still does say some of those things. And you can still get it to those things. And by the way, I've gotten emails from parents since we gave that presentation, and their kids have been independently found doing things like that that.

Kara Swisher: Doing things like that.

Tristan Harris: Yeah.

Kara Swisher: So they just can't anticipate all the problems.

Tristan Harris: Well, it's actually worse than that. It's just important for listeners to know. Just to be fair to Snapchat, they actually did not roll that My AI bot out to all of its, I can't remember if it's 700 million users. They didn't roll it out to all their users. They rolled it out to only paid subscribers at first, which is something like two to three million users. But of course just two weeks ago or something like that, they released it to all their users. Why did they do that? Because they're in a race to dominate that intimate spot in your life. Everyone wants to be the Scarlett Johansson Her AI bot in your ear.

Kara Swisher: You both signed a letter calling for the six-month pause on giant AI experiments. Elon did too. Elon Musk did too.

Tristan Harris: It's unfortunate that that letter got defined by Elon's participation in it.

Kara Swisher: Yes, because he looked like he was doing his own business.

Tristan Harris: Well, later obviously he then also started his own AI company. And so obviously it de-legitimizes...

Kara Swisher: Yeah. He also laughed and said he knew it would be futile to sign it. So why make that? Many people think it was a futile effort.

Tristan Harris: Well, these are separate topics. I want to make sure we really slow down and actually distinguish here.

Kara Swisher: Okay.

Tristan Harris: The founders of the field of machine learning helped sign that letter. Steve Wozniak started the letter, the co-founder of Siri, signed the letter. Andrew Yang, et cetera, all of us at Center for Humane Technology. That letter is because the Overton window of society about how unsafe and dangerous this is, was not well known. The purpose of that letter was to make it very well known that this field is much more dangerous than what people understand. And we know the Future of Life Institute folks who were really kind of spearheading the letter. There was a lot of debate about what is the appropriate time to call for a slowdown. And by the way, I think slowdown is also badly named on retrospect. I think something like redirection of all the energy of those labs into safety work and safety research and guardrails. So imagine it's six months instead of an AI winter, an AI harvest, an AI summer, where you harvest the benefits that you have, you do understanding on what are the capabilities inside of everything that's been released.

Kara Swisher: Did you imagine this was going to happen? That they were, would go, "Oh yes. Oh, yes. I see your point, sir."

Tristan Harris: Well, connected to the team that did it and kind of being privy to some of the internal conversations, I think we were all surprised how many incredible people did sign the letter.

Kara Swisher: They did. Yeah.

Tristan Harris: Many people signed the letter. It's funny that people look at it and maybe say, "This is futile." But it's like saying just because something is hard doesn't mean it shouldn't be the intention. And one of the interesting things is that if you talk to an engineer and you say, "Oh, we're going to build this AGI thing", and they're like, "Oh, that sounds really hard." But it's like, we're so compelled by the idea of building these AI systems, these AGI systems that God that I could talk to that they say, "I don't care how hard it is." And so they keep racing towards it. And it's been a hundred years that, whatever, 50 years that people have been working on this. In other words, we don't say because something's hard, "Oh we shouldn't keep going and try to build it anyway." Whereas if I say, "Coordination is hard for the whole world", people say, "Oh, let's just throw up our hands and say it's never going to happen."

Kara Swisher: Right.

We need to get good at coordination. All of our world's problems are coordination problems.

Right. We do it with nuclear energy, we do it with a lot of things.

Tristan Harris: We have limited nukes to nine countries. Just to put a pin on it though. If I said, "It's inevitable that all countries are going to get nukes, let's not do anything about it. In fact, let's just let every country pursue it and not do anything", we probably wouldn't be here today. A lot of people had to be very concerned about it and move into action to say something different needs to happen.

Kara Swisher: But a nuclear war we got. We saw it. It happened with the atom bomb.

Tristan Harris: Yeah.

Kara Swisher: So give me your best case against a pause. And one of the more compelling criticisms is US is going to fall behind China. This is something I heard from Mark Zuckerberg about social media in general, or tech in general.

Tristan Harris: Which is interesting because I would argue...

Kara Swisher: China though, they use the same Xi or me argument every time. They drag it out. But it's concerning. It is. It absolutely is. China has shown itself to have very few governance on itself.

Tristan Harris: I would say unregulated deployment of AI would be the reason we lose to China. If worse actors do beat you in dominance, in deploying AI, people with no morals, with no safety considerations, with no concerns, with different values as a future of the world kind of society, Chinese digital authoritarianism values or something like that, or Chinese Communist Party values, then we certainly won't want to lose to that. So I think if there was a sincere risk that that would happen, there would be a good reason to say let's not call for that.

Kara Swisher: Okay. All right.

Tristan Harris: But I would actually argue that the unregulated deployment of AI is what is causing the West to lose to China. Let me give you the example of social media. Social media was the unregulated deployment of AI to society. The breakdown of democracy's ability to coordinate, because we no longer have insured...

Kara Swisher: That's good for authoritarianism.

Tristan Harris: That's really good for the authoritarianism. Why are democracies backsliding everywhere around the world, all at once? Barbara F. Walter wrote a book called How The Next Civil Wars Start. She talks about democracies, democracies that are backsliding everywhere. I'm not blaming it all on social media, but we're seeing it happen rapidly in all these countries that have been governed by the information environment created by social media. And if a society cannot coordinate, can it deal with poverty? Can it deal with inequality? Can it deal with climate change?

Kara Swisher: So we shot ourself in the foot and now we're going for the arms.

Tristan Harris: Yeah, exactly.

Kara Swisher: Right. That kind of thing.

I've interviewed you a number of times, one we did in 2017, as I said, before you and Aza founded the Center for Humane Technology. Back then you were focused on social media as we discussed earlier, showing why revenue models built on monetizing our attention are bad for us. Because a lot of this is about monetization and who's going to have the next intimate relationship; which they've been trying to do forever in different ways, through Siri and all kinds of different things. But now they really want you to be theirs essentially. Let's play a clip from it.

Tristan Harris: Right now, essentially Apple, Google and Facebook are kind of like these private companies who collectively are the urban planners of a billion people's attentional landscape.

Kara Swisher: That's a great way to put it.

Tristan Harris: We kind of all live in this invisible city.

Kara Swisher: Right. Which they created.

Tristan Harris: Which they created. And the question, unlike a democracy where you have some civic representation and you can say, "Well, who's the mayor? And should there be a stoplight there? Stoplight on our phone or blinker signals between the cars", or these kinds of things. We don't have any representation except, if we don't use the product or don't buy it. And that's not really representation because the city itself is...

Kara Swisher: So attention taxation without representation.

Tristan Harris: Maybe, yeah. So I think there's this question of how do we create that accountability loop.

Kara Swisher: That was very well put. And now we took it further. I said, "It's like the purge." They actually own the city and they don't do anything.

Tristan Harris: Oh yeah.

Kara Swisher: We can't do anything and they won't do anything. They have no stop signs, they have no streets, they have no sewage, or everything else. So I took your thought a step further. Talk about AI firms becoming the new urban planners of the, I guess attentional landscape. Because that's what they want. It's more than attention they want. They want to own you, right?

Tristan Harris: Yeah.

Kara Swisher: I mean, is what you're saying.

Tristan Harris: Well, I want to separate between two different economies. So there's the engagement economy, which is the race to dominate, own, and commodify human experience. So that's the social media.

Kara Swisher: Social media.

Tristan Harris: Social media is the biggest player in that space.

Kara Swisher: Right.

Tristan Harris: But VR is in that space. YouTube is in that space. Netflix is in that space. It's the race to say...

Kara Swisher: Look at me.

Tristan Harris: Look at me. All the things that construct your reality, that determine from the moment you wake up and your eyes open to the moment your eyes close at the end of the night, who owns that space?

Kara Swisher: Your attention.

Tristan Harris: That's the engagement economy. That's the attention economy. And there's specific actors in that space. AI will be applied to that economy, just like AI will be applied to all sorts of other economies also. The cyber hacking economy; it will be applied to the battery storage.

Kara Swisher: It's more like the internet.

Tristan Harris: Yeah. AI is a much bigger thing. So there's a sub part of the AI economy, which is the engagement economy. And AI will supercharge the harms of social media there. Because before we had people AB testing a handful of messages on social media and figuring out, like Cambridge Analytica, which one works best for each political tribe. Now you're going to have AIs that do that, and there's a paper out called, I think it's called silicon sampling. So you can actually sample a virtual group. Instead of running Frank Luntz focus groups around the world, you can kind of have a language bot, chatbot, that you talk to that and it will answer questions as if its someone that's a 35-year-old in Kansas City, has two kids. And so you can run even perfect message testing.

Kara Swisher: Right. So you don't need to talk to people.

Tristan Harris: So you don't need to talk to people anymore.

Kara Swisher: You know what they're going to say.

Tristan Harris: Yeah. You can do a million things like that. And so the loneliness crisis that we see, the mental health crisis that we see, the sexualization of young kids that we see, the online harassment situation that we see, all that's just going to get supercharged with AI. And the ability to create alpha persuade, which is just like there was AlphaGo and alpha chess, where the system's playing chess against itself and kind of getting much better. It's now going to be able to hyper-manipulate you and hyper-persuade you.

Kara Swisher: So what you're talking about is social media as a lower being than AI. AI powers everything. Social media is one, but we couldn't even regulate social media. Is society aware of the need for regulation, since we didn't do it for social media?

Tristan Harris: So the point we made in this AI Dilemma presentation is that we were too late with social media because we waited for it to entangle itself with journalism, with media, with elections, with business. Because now businesses can only reach their consumers if they have an Instagram page and use marketing on Facebook and Instagram and so on. Social media captured too many of the fundamental life organs of how our society works. And that's why it's been very hard to regulate. I mean, certain parties benefit, certain politicians benefit. Would you want to ban TikTok if you're a politician or a party that's currently winning a lot of elections by being really good at TikTok?

Kara Swisher: Right.

Tristan Harris: So once things start to entangle themselves, it's very hard to regulate them. There's too many invested interests. With AI, we have not yet allowed this thing to roll out. I mean now it's obviously happening incredibly fast. When we gave the presentation a few months ago, the whole point of it was, before GPT-4, we need to act before this happens. One good example of this happening in history was a treaty to ban blinding laser weapons from the battlefield before they were actually ever used.

Kara Swisher: To blind the soldiers.

Tristan Harris: To blind soldiers, yes. This would be a high energy laser that has the capability. Point it at them, and it just blinds them. But we're just like, "You know what? In the game of war, which is a ruthless game where you kill other human beings, even as ruthless as that game is, we don't want to allow that." And even before it was ever deployed, that was one of maybe the most optimistic examples where humanity could sort of use our higher selves to recognize that future gain...

Kara Swisher: Goes into the killer robot part of the portion of the show, right?

Tristan Harris: Right. Then there's the slaughter bots. How do we ban autonomous weapons? How do we ban recombinant DNA engineering and human cloning, things like this. And so this is another one of those situations. And we need to look to, especially the example of the blinding laser weapons, because that was an advance of the technology ever getting fully deployed. Because a lot of the guardrails that we're going to need internationally are going to be saying no one would want that future race to happen. So let's prevent that race.

Kara Swisher: Right. But that's nation states. Now AI, anybody could do it. The same thing with CRISPR though. They definitely, the scientists got together and had standards. And this is much easier to be able to do what you want if we are all in a group together coordinating this.

Tristan Harris: So if I want to steelman the AI doomers and the P-doomers that have a really high number for that P(doom) number, it's because it's so hard to prevent the proliferation that many people think that we're doomed. Just to really clear in why that's also a very legitimate thing.

Kara Swisher: That would be my biggest P(doom). This is too easy for a lot of people.

Tristan Harris: It's too easy. So let's just hang there for a moment. Just really recognize that. That's not being a doomer, that's just being an honest viewer of these are the risks. Now, if something other were to happen, you could involve governments and law to say, "Hey, we need to get maybe more restrictive about GitHub and Hugging Face and where these models go. Maybe we need export controls." Just like there's 3D printed guns as a file, you can't just send those around the open internet. We put export controls on those kinds of things. It's a dangerous kind of information.

So now imagine there's a new kind of information that's not a 3D printing gun, but it's like a 3D printing gun that actually self replicates and self improves and gets into a bigger gun.

Kara Swisher: And it builds itself.

Tristan Harris: And builds itself. That's a new class. That's not just free speech. The founding fathers couldn't anticipate something that self replicates and self improves being a class of speech. That's not the kind of speech that they were trying to protect.

Kara Swisher: Right.

Tristan Harris: Part of what we need here are new legal categories for these new kinds of speech.

Kara Swisher: Sam Altman, who runs OpenAI, was on the Hill calling for AI regulation. They all are. You can't say you didn't warn them, right? A lot of tech CEOs have claimed they want regulation, but they've also spent a lot of money previously on stopping antitrust, stopping algorithm transparency, stopping any privacy regulation. Do you believe this class of CEOs? Because a lot of them are saying, "This is dangerous. Would you please regulate this?"

Tristan Harris: Yeah. So you're pointing to what happened with social media, which was that publicly they would say, "We need regulation, we need regulation, we need regulation." When you talk to the staffers...

Kara Swisher: They never said this is dangerous.

Tristan Harris: They never said dangerous.

Kara Swisher: He says dangerous.

Tristan Harris: He says dangerous. And I want a golf clap that we always want to endorse and celebrate when there is actually an honest recognition of the risks. I mean to Sam Altman's credit, he has been saying in public settings, I think much to the chagrin and maybe his investors and other folks, that there are existential risks here. I mean what CEO goes out there saying, "This could actually wipe out humanity," and not just because of jobs. So we should celebrate that he's being honest about the risks because we actually do need an honest conversation about it.

However, as you said in the history of social media, it is very easy to publicly advocate for regulation, and then your policy teams follow up with all the staffers and then say, "let me redline this redline that. That's never going to work." And they just sort of stall it so that nothing actually ever happens.

Kara Swisher: Right.

Tristan Harris: I don't think it's that bad faith in this context. I do think that some kind of regulation's needed. Sam Altman talked about GPU licensing. Licensing doing a training run. If you're going to run a large frontier model, you're going to do a massive training run. You got a license to do that. Just like we had the Wuhan Institute of Virology was a biosafety level four lab doing advanced gain of function research. If you're building a level four lab, you need level four practices and responsibilities. Even there though, we know that that may not have been enough, whatever safety practices.

Kara Swisher: Right.

Tristan Harris: We're now building AI systems that are super advanced and the question is, do we actually have the safety practices?

Kara Swisher: Are we treating it like a top lab?

Tristan Harris: Well, the first thing is are we treating it that way? And then the second is, do we even know what would constitute safety? This is getting the end question you're asking, can we even do this safely? Is that even possible?

Kara Swisher: Right.

Tristan Harris: Because think of AI as like a biosafety level 10 lab. Imagine we had something called, I'm inventing it right now, but a biosafety level 10 lab where I invent a pathogen that the second is released, it kills everyone instantly. Let's just imagine that that was actually possible. Well, you might say, "Well, let's let people have that scientific capacity. We want to just see is that even possible. We want to test it so we maybe can build a vaccine or prevention systems against a pathogen that could kill everyone instantly." But the question is to do that experimental research, what if we didn't have biosafety level 10 practices. We only had biosafety level 10 dangerous capabilities. Would we want to pursue biosafety level 10 labs?

I think that AI, the deeper question is, you cannot have the power of Gods without the wisdom, love, and prudence of Gods. And right now we are handing out and democratizing God-like powers without actually even knowing what would constitute the love, prudence and wisdom that's needed for them. And I think the story and the parable of Lord of the Rings is, why do they want to throw the ring into Mount Doom? There's some kinds of powers that when you see them, you say, "If we're not actually wise enough to hold this ring and put it on", we have to know which rings we have to say, "Hey, let's collectively not put on that ring."

Kara Swisher: Right. I get that. I understand that. One of the things is that when you get this dramatic, like I said at the beginning, does that push people off? Like this is a pathogen we get. We've just been through COVID and that was bad enough. And there's probably a pathogen that could kill people instantly. It's not how people think.

Tristan Harris: Yeah. Well let's actually just make that example real for a second, because that was a hypothetical thing of about safety level 10 thing.

Kara Swisher: Right.

Tristan Harris: Can AI accelerate the development of pathogens and gain of function research and people tinkering with dangerous lethal bioweapons? Can it democratize that? Can it make more people able to do that? More people be able to make household explosives with household materials? Yes. We don't want that. That's really dangerous. That's a very concrete thing. That's not AI doomers. There's real concrete stuff we have to respond to here.

Kara Swisher: Okay.

We will be back in a minute.

Tell me something that AI could be good for, because I talk about that. Because I think I'm a little less extreme than you. There are, and I think at the beginning of the internet I was like, "This could be great." And of course then you saw them not worrying about the not so great. And I think it's sort of that tools and weapons, speaking of which from Microsoft, that was the Microsoft president Brad Smith talked about tools and weapons. A knife is a tool and a weapon. So what is the tool part of this that is a good thing?

Tristan Harris: So first of all, I think this is one of those things, just like we say is the AI sentient, that when people hear me saying all this, they think I don't hear or don't know about or aren't talking about all the positives it can do. This is another fallacy of how human brains work.

Kara Swisher: Yeah.

Tristan Harris: Just like we get obsessed with the question of, is it sentient, we get obsessed with the onesidedness of... it has all the positives. Just as fast as you can design cyber weapons with AI and accelerate the creation of that, you can also identify all the vulnerabilities in code or many vulnerabilities in code. You can invent cures to diseases. You can invent new solutions for battery storage. As I said in The Social Dilemma, what's going to be confusing about this era is its simultaneous utopia and dystopia.

Kara Swisher: I can't think of so many good things about social media. I couldn't. I can think of dozens here. Dozens here. And there I was like, "Maybe we'll all get along and do better."

Tristan Harris: Social media is increasing the flows of information. People are able to maintain many more relationships; old high school sweethearts.

Kara Swisher: Not like this. This is gene folding, this is drug discovery. This is real movement forward.

Tristan Harris: Yeah, absolutely.

Kara Swisher: Right.

Tristan Harris: I'll tell a story. So the real confusing thing is, is it possible on the current development path to get those goods without the bads? What if it was not possible? What if I can only get that the synthetic biology capabilities that let me solve problems, but there was no way to do it without also enabling bad guys.

Kara Swisher: Than to create this pathogen that you're talking about, for example.

Tristan Harris: So just to make it personal, my mother died of cancer. And I like any human being would do anything to have my mother still be here with me. And if you told me that there was an AI that was going to be able to discover a cure for my mother that would have her still be with me today, obviously I would want that cure. If you told me that the only way for that cure to be developed was to also unleash capabilities that the world would get wrecked.

Kara Swisher: This is a dinner party, one of those dinner party questions. Would you kill a hundred million people to save?

Tristan Harris: But it's real.

Kara Swisher: Yeah.

Tristan Harris: I'm just saying there's certain domains where there's no way to do the one side without doing the other side.

Kara Swisher: Right.

Tristan Harris: And if you told me that, just really on a personal level, as much as I want my mom to be here today, I would not have made that trade.

Kara Swisher: Well you're talking about an old Paul Virilio quote, which is, "You can't have a ship without a shipwreck or electricity without the electric chair." We do that every day. Net cars have been great, net they've been bad now. You know what I mean?

Tristan Harris: But if you have godlike powers that can kind of break society in much more fundamental ways. So now again, we're talking about benefits that are literally god-like invented solutions for every problem, but if it also just undermines the existence of how life will work...

Kara Swisher: So that's your greatest worry is this idea of reality fracturing in ways that are impossible to get back.

Tristan Harris: No. I mean all of it together. If AI is unleashed and democratized to everybody, no matter how high the tower of benefits that AI assembles, if it also simultaneously crumbles the foundation of that tower, it won't really matter. What kind of society can receive a cancer drug if no one knows what's true, there's cyber attacks everywhere, things are blowing up and there's pathogens that have locked down the world. Again, think about how bad COVID was. People forget. Going through one pandemic, just one pandemic. Imagine if that just happens a few more times. We saw the edges of our supply chains. We saw how much money had to be printed to keep the economy going. It's pretty easy to break society if you have a few more of these things going. And so again, how will cancer drugs sort of flow in that society that has kind of stopped working? And I don't mean, again, AI doom, Eliezer Yudkowsky AGI kills everybody in one instant. I'm talking about dysfunction at a scale that is so much greater.

Kara Swisher: Are we getting closer to regulation? Did you find those hearings... Did you have any good takeaways from them, and where is it going to go from here?

Tristan Harris: Who knows where it's going to go. I didn't see all of the hearing. I was happy to see a couple things, which is based on structural issues. So one was actually the repeated discussion of multilateral bodies. So something like an IAEA, like the International Atomic Energy Agency, but something like that for AI, that's actually doing global monitoring and regulation of AI, systems of large frontier AI systems. I think Sam was proposing that; that was repeated several times. I was surprised to see that. I think that's actually great because it is a global problem. What's the answer when we develop nuclear weapons? Is it that Congress passes a law to deal with nukes here? No. It's a global coordination around how do we limit nukes to nine countries. How do we make sure we don't do above ground nuclear testing? So I was happy to see that in the hearing.

I was also happy to see multiple members of Congress, including I think it was Lindsey Graham, and Republicans who are typically not for new regulatory agencies, them saying they recognize that we need one. E.O. Wilson; if we have paleolithic emotions, medieval institutions, and God-like tech, medieval institutions and medieval laws, 18th century ideas, 19th century laws and ideas, don't match for 21st century issues like replicant speech. Larry Lessig has a paper out about replicant speech. Should we protect the speech of generative robots the same way we protect free speech, that the founding fathers had totally different ideas what that was about? No, we need to update those laws. Part of our medieval institutions are institutions that don't move as fast as the godlike tech. So if a virus is moving at 21st century speeds and your immune system is moving at 18th century speeds, your immune system is being deregulation.

Kara Swisher: So do you have any hope for any significant legislation? I mean, Vice President Harris met with a... They're all meeting with everybody for sure. And early, compared to the other things.

Tristan Harris: I don't remember Kara, but we did that briefing in DC back here in whatever it was, February or March, we said one of the things we really want to happen is for the White House to convene a gathering of all the CEOs. And that, I would've never thought would've ever happened. And it did happen.

Kara Swisher: Yes.

Tristan Harris: I would've never thought that there would be a hearing.

Kara Swisher: And they mentioned at the G7 this week.

Tristan Harris: And they did it. They mentioned the G7 this week. So there's things that are moving. And I want people to be optimistic, by the way. There needs to be a massive effort and coordinated response to make the right things happen here.

Kara Swisher: Right. Vice President Harris led that meeting and told them they have ethical, moral and legal responsibility to ensure the safety and security of their products. They certainly don't seem protected by section 230. They're probably not protected. There is liability attached to some of this, which could be good that.

Tristan Harris: That's good. We talked to people inside the company at this point. Because all we're trying to do is figure out what needs to happen. And often the people inside the companies who work on safety teams will say, "I can't advocate for this publicly, but we need liability." Because talking about responsibility and ethics, just get bulldozed by incentives. There needs to be liability that creates real guardrails.

Kara Swisher: Right.

Let's do a lightning round. What you would say to the following people if they were here right now. Sam Altman, CEO of OpenAI, what would you say to him, Tristan?

Tristan Harris: Gather all of the top leaders to negotiate a coordinated way to get this right. Move at a pace that we can get this right, including working with the Chinese and getting multilateral negotiations happening. And say that's what needs to happen. It's not about what you do with your company and your safety practices and how much...

Kara Swisher: Multilateral.

Tristan Harris: Get coordination.

Kara Swisher: Satya Nadella and Sundar Pichai; going to mush them together.

Tristan Harris: Retract the arms race. Instead of saying, "Let's make Google dance", which is what Satya Nadella said. We have to find a way to move back into a domain of advanced capabilities being held back. Buying ourselves a little bit more time matters.

Kara Swisher: Yeah. Well they've been sick of being pantsed the entire last decade. I think they want to do that in some fashion.

Tristan Harris: Understood.

Kara Swisher: Reid Hoffman, Mustafa Suleyman of Inflection AI, which put out a chatbot this month.

Tristan Harris: I mean, honestly, it would be the same things with Sam. It's like, everyone needs to work together to get this right.

Kara Swisher: Okay.

Tristan Harris: We need to see this as dangerous for all of humanity. This isn't us versus the tech companies. This is, all of us are human beings and there's dangerous outcomes that land for all of us.

Kara Swisher: What about Elon Musk? He signed the AI pause letter, has been outspoken on the danger for years. He was one of the earliest people that were talking about it, along with Sam as I recall, a decade ago. But he of course started his own company, X AI. When he wants to get to the truth of AI, whatever that means.

Tristan Harris: We need to escape this logic of, "I don't think the other guys are going to do it right, so I'm going to therefore start my own thing to do it safely." Which is how we got to the arms race that's now driving all the unsafety. And so the logic of, "I don't believe in the way the other guys are doing it, and mostly for competitive reasons probably underneath the hood, I'm doing my own thing", that logic doesn't work.

Kara Swisher: Yeah, he's very competitive. Do you blame them personally for putting us at risk? Or is it just one of these group things that everyone goes along?

Tristan Harris: There's this really interesting dynamic where when there is a race; which, all the problems are driven by races. If I don't do the mining in that version, place, or if I don't do the deforestation, I just lose to the guy that will.

Kara Swisher: Right.

Tristan Harris: If I don't dump the chemicals in and my competitors do...

Kara Swisher: And I'll do more safely.

Tristan Harris: And I'll do it more safely. So better me doing it than the other guy as long as I get my profit. And so everyone has that self-reinforcing logic. So there's races everywhere that are the real driver of most of the issues that we're seeing. And there's a temptation once we diagnose it as a race, a bad race, to then absolve the companies of responsibility. I think we have to do both. There's both a race and also, Satya Nadella and Sam helped accelerate that race in a way that actually, we weren't trajectoring that way. There was human choices involved at that moment in the timeline. I talked to people who helped found some of the original AGI labs early in the day. They said, if we go back 15 years, they would've said, "Let's put a ban on pursuing artificial general intelligence, building these large systems that ingest the world's knowledge about everything. We don't need to do that. We should be building advanced applied AI systems, like alpha fold that says let's do specific targeted research domains and applications." If we were living in that world, how different might we be?

We have three rules of technology we put in that AI Dilemma presentation. When you invent a new technology, you create a new class of responsibilities. Second rule of technology. If the new technology you invent confers power, it will start a race. If I don't adopt the plow and start out-competing the other society, I'll lose to the guy that does adopt the plow. I don't adopt social media to get more efficient about yeah; et cetera. So it starts a race. Third rule of technology. If you do not coordinate that race, the race will end in tragedy. We need to become a society that is incredibly good at identifying bad games rather than bad guys. Right now, all we do have bad guys. We have, again, CEOs that do bear some responsibility for some choices. But right now that drives up polarization, because you put all the energy into going after one CEO or one company, when we have to get good at slaying bad games.

Kara Swisher: Well, except wouldn't you agree that one of the reasons social media got so out of whack was because of Mark Zuckerberg and his huge power? He had a power over the most biggest thing, and just was both badly educated.

Tristan Harris: Mark made a ton of bad decisions while denying many of the harms most of the way through until just recently.

Kara Swisher: Yeah.

Tristan Harris: Including that it was a crazy idea that fake news had anything to do with the election. Later they found the Russia stuff was, "Oh, this is all overblown." Which, yeah, I understand. There's the Trump-Russia stuff, there may have been overblown stuff there, but the Facebook content, they said, "Oh, it didn't really reach that many people." And it ended up reaching 150 million Americans.

Kara Swisher: No, I get it.

Tristan Harris: Facebook's own research said that 64% of extreme...

Kara Swisher: I've sat on the other side of the fence.

Tristan Harris: We can go on for forever about that.

Kara Swisher: Geoffrey Hinton, who was known as one of the godfathers of AI, not the only one, had recently been sounding the alarm. Do you think others would follow suit? That was a big deal when he did that.

Tristan Harris: It really was.

Kara Swisher: I was very aware of him and AI. Do you think it'll change the direction or is he just Robert Oppenheimer saying, "I have become death?"

Tristan Harris: One of the things that struck me both, I came out too, right? I was an early person coming out and I've seen the effects of insiders coming out. Francis Haugen, the Facebook whistleblower's a close friend of mine. And her coming out made a really big difference. The Social Dilemma I know impacted her. It legitimized for many people inside the companies that they felt like something was wrong, and now many more people came out. I think the more people come out, the more the big names come out, the Geoff Hintons come out. It actually makes more people question. I think this few days ago, there's now a street protest outside of DeepMind's headquarters in London saying, "We need to pause AI." I don't know if you saw that.

Kara Swisher: No. I see it's comparable to climate change in a lot of ways.

Tristan Harris: There are real people inside their own companies that are saying, "There's a problem here." Which is why it's really important that when the people who are making something who know it most intimately are saying there's a real problem here, when the head product guy at Twitter says, "I don't let my own kids use social media." That's all you need to know about whether something is good or safe.

Kara Swisher: So one of the things, there's some proposals you brought up, there's one based on a work by Taiwan's digital minister, who's so creative, where a hundred regular people get in a room with AI experts and they come out with a proposal.

Tristan Harris: Mm-hmm.

Kara Swisher: That's an interesting one. You come up with one having a national televised discussion, major AI labs, lead safety experts, and other civic actors talk on TV. That's hard because on one hand I could see that working, but not working.

Tristan Harris: Yeah it has to be done carefully. Let me explain the Taiwan one really quickly.

Kara Swisher: Okay.

Tristan Harris: Okay. So let's imagine there's kind of two attractors for where the world is going right now. One attractor is, I trust everyone to do the right thing and I'm going to distribute god-like AI powers, superhuman powers to everyone. Everyone can build bioweapons, everyone can make generative media, find loopholes in law, manipulate religions, do fake everything. That world lands in continual chaos and catastrophe, because it's just basically I'm handing everyone out the power to do anything.

Kara Swisher: Yeah, totally. Yeah. Everyone had superpowers, yeah.

Tristan Harris: Right. So that's one outcome. That's one attractor. Think of it like a 3D field, and it's like sucking the world into one gravity while it's just continuing to catastrophe.

Kara Swisher: Kind of like guns, but go ahead.

Tristan Harris: Yeah.

The other side is dystopia, which is instead of trusting everyone to do the right thing with these superhuman powers, I don't trust anyone to do the right thing. So I create this dystopian state that has surveillance and monitors everyone. That's kind of the Chinese digital authoritarianism outcome. That's the other deep attractor for the world, given this new tech that's entering into the world. So the world's currently moving towards both of those. And actually the more frequently the continual catastrophes happen, the more it's going to drive us towards the direction of the dystopia. So in both cases, we're getting a self-reinforcing loop.

So the reason I mentioned Taiwan is what we need is a middle way or third attractor, which is what has the values of an open society, a democratic society, in which people have freedom but instead of naively trusting everyone to do the right thing, instead of also not trusting anyone to do the right thing, we have what's called warranted trust. So think of it as a loop. Technology, to the degree it impacts society, has to constitute a wiser, more responsible, more enlightened culture. A more enlightened culture supports stronger upgraded institutions. Those upgraded institutions sets the right kind of regulatory or guardrails, et cetera, for better technology that then is in a loop with constituting better culture. That's the upward spiral.

We are currently living in the downward spiral. Technology decoheres, digs outrage, lonelifies culture. That incoherent culture can't support any institutional responses to anything. That incapacitated, dysfunctional set of institutions doesn't regulate technology, which allows the downward spiral to continue. The upward spiral is what we need to get to.

And the third way, what Taiwan is doing is actually proving that you can use technology in a way that gets you the upward spiral. Audrey Tang's work is showing that you can use AI to find unlikely consensus across groups. There's only so many people that can fit into that town hall and get mad at each other. What if she creates a digitally augmented process where people put in all their ideas and opinions about AI and we can actually use AI to find the coherence, the shared areas of agreement that we all share, and do that even faster than we could do without the tech.

Kara Swisher: Right.

Tristan Harris: So this is not techno utopianism, it's techno realism, of applying the AI to get a faster OODA loop, a faster observe, orient, decide, and act loop, so that the institutions are moving as fast as the evolutionary pace of technology. And she's got the best, closest example to that. And that's kind of part of what a third attractor needs to identify.

Kara Swisher: Where people feel that they've been put in and at the same time don't feel the need to scream, which is absolutely true. She's really quite something. Having a national debate about it, I think people just take away whatever they want from it.

Tristan Harris: Yeah. Let me explain that though, which was modeled after the film The Day After. So in the previous era of a new technology that had the power to...

Kara Swisher: I remember. I was there in college when that happened.

Tristan Harris: In college when it came out?

Kara Swisher: Mm-hmm.

Tristan Harris: I was not born yet.

Kara Swisher: Let me just explain. This is a movie about the nuclear bomb blowing up and they convened groups all over the country to talk about it; watch the movie, and then discuss it. And it really was terrifying at the time. But we were all joined together in a way we're not anymore. I can't even imagine that happening right now.

Tristan Harris: It was a made for TV movie commissioned by ABC where the director, Nicholas Meyer, who also directed Star Trek Two: The Wrath of Khan and some other great films, they put together this film that was basically noticing that nuclear war, the possibility of it, existed in a repressed place inside the human mind. No one wanted to think about this thing that was ever present. That actually was a real possibility, because it was the active Cold War, and it was increasing and escalating with Reagan and Gorbachev. So they decided, let's make a film that became the largest made for TV watched film in all of TV history. A hundred million Americans tuned in, I think it was 1983, watched it once.

They had a whole PR campaign; put your kids to bed early, which actually increased the number of people who actually didn't watch it with their kids. Reagan's biographer later, several years later said that Reagan got depressed for weeks. He watched in the White House film studio. And when the Reykjavik Accords happened, because they actually, I should mentioned, they aired the film the day after in the Soviet Union a few years later in 1987. And it scared basically the bejesus out of both the Russians and the US.

Kara Swisher: Yeah. It was quite something at the time.

Tristan Harris: And it made visible and visceral the repressed idea of what we were actually facing. We actually had the power to destroy ourselves. And it made that visible and visceral for the first time. And the important point that we mentioned this AI Dilemma talk that we put online, is that after this one and a half hour, whatever it was film, they aired a one-hour debate where they had Carl Sagan and Henry Kissinger and Brent Scowcroft, and Elie Wiesel studied the Holocaust, to really debate what we were facing. And that was a democratic way of saying, "We don't want five people at the Department of Defense in Russia and the US deciding whether humanity exists tomorrow or not."

Kara Swisher: Yeah.

Tristan Harris: And similarly, I think we need that kind of debate. So that's the idea. I don't know about a TV broadcasting.

Kara Swisher: Well, I don't think it'll work today, honestly. I don't. What's interesting is that was very effective. That's an interesting thing to talk about, The Day After, because it did scare the bejesus. Watching Jason Robards disintegrate in real time was disturbing.

Tristan Harris: Yeah.

Kara Swisher: But there was nothing like that. And now there is a lot like that. Everybody is constantly hit with information every day. It was unique because we used to have a commonality that we don't have.

So you have gone on Glenn Beck podcast, God save you, Brian Kilmeade podcast.

Tristan Harris: We do a lot of media across the board.

Kara Swisher: Right, exactly.

Do they react differently from your message than progressive audiences?

Tristan Harris: No.

Kara Swisher: Because again, can they split? Progressive, tech companies are bad.

Tristan Harris: Well, let me say it differently.

Kara Swisher: And conservatives, surveillance and the deep state.

Tristan Harris: Yeah, well, exactly. Social media got polarized. So it's actually one of the reasons I'm doing a lot of media across the spectrum, is I have a deep fear that this will get unnecessarily politicized. That would be the worst thing to have happen, is when there's deep risks for everybody, it does not matter which political beliefs you hold. This really should bring us together. And so I try to do media across the spectrum so that we can get universal consensus, that this is a risk to everyone and everything, and the values that we have and people's ability to live in the future that we care about. I do this because I really want to live in a future that kids can be raised and we can live in a good world as best as we can. We're facing a lot of dark outcomes.

Kara Swisher: Right.

Tristan Harris: There's a spectrum of those dark outcomes. Let's live on the lighter side of that spectrum rather than the darkest side, where maybe the lights go out.

Kara Swisher: I have one last question. How do you think the media has been covering it? Because there is a pressure, if you cover it too negatively, it's like, "Oh, come on. Don't you see the better? Are you missing the bigger picture?" And I know from my personal experience, I'm so sick of being called the bummer bite or an irritant. It gets exhausting.

Tristan Harris: Yeah.

Kara Swisher: But at the same time, you do want to see, maybe this time we can do it better. Give me hope here, because I definitely feel the pressure not to be so negative. And I still am, I don't care. And I think in the end, both of us, were right back then, but it doesn't feel good being right.

Tristan Harris: Everything creates externalities, effects that show up on other people's balance sheets. If you are a doomer and you think you're just communicating honestly, but you end up terrifying people, maybe some shooters come around and they start doing violent things because they've been terrorized by what you've shared. I think about that a lot. I think a lot about responsible communication.

So I think there's a really important thing here, which is that there's kind of three psychological places that I think people are landing. The first is what we call pre-tragic. I borrow this from a mentor, Daniel Schmachtenberger, who we've done the Joe Rogan show with. Pre-tragic is someone who actually doesn't want to look at the tragedy, whether it's climate or some of the AI issues that are facing us or social media having downsides. Any issue where there is a tragedy, but we don't want to metabolize the tragedy, so we stay in naive optimism, they call this kind of person a pre-tragic person. Because there's a kind of denial and repression of actual honest things that are facing us.

Kara Swisher: Right.

Tristan Harris: Because I want to believe, "Well, things always work out in the end. Humanity always figures it out. We muddle our way through." Those things are partially true too, but let's be really clear about the rest. Okay, so that's the pre-tragic.

Then there's the person who then stares at the tragedy, and then people tend to get stuck in tragedy. You either get depressed or you become nihilistic. Or the other thing that can happen, is it's too hard and you bounce back into pre-tragic. You bounce back into, "I'm just ignore that information, go back to my optimism", because it's just too hard to sit in the tragedy.

There's a third place to go, which is we call post-tragic, where you actually stare face to face with the actual constraints that are facing us, which actually means accepting and grieving through some of the realities that we are facing. I've done that work personally, and it's not about me. I just mean that I think it's a very hard thing to do. It's the humanities' rite of passage. You have to go through the dark night of the soul and be with that so you can be with the actual dimensions of the problems that we're dealing with. Because then when you do solutions on the other side of that, when you're thinking about what do we do, now you're honest about the space. You're honest about what it would take to do something about it.

Kara Swisher: So you're not negative.

Tristan Harris: No.

Kara Swisher: But people will cast you as that. You don't feel you are.

Tristan Harris: So there's something called pre-transfallacy where someone who's post-tragic can sound like someone on the other side.

Kara Swisher: Yes.

Tristan Harris: It can sound confusing. So I can sound like a doomer, but really it is, I'm trying to communicate clearly. People often ask me, "Am I an optimist?"

Kara Swisher: Are you a prepper?

Tristan Harris: No. No.

Kara Swisher: Had to ask, had to ask. Sam Altman has his little home.

Tristan Harris: I know he does. I know he does. Yeah.

Kara Swisher: He once asked me, "What was my plan?" Just joking, we were joking around about it, I said, "Well, you're smaller than I. I'm going to beat you up and take your things and take your whole plan." He's like, "That's a good plan."

Tristan Harris: That's like a thousand, or whatever it is.

Kara Swisher: Yeah. He goes, "That's a good plan." I go, "It's an excellent plan."

Tristan Harris: Yeah.

Kara Swisher: "I think I can take you", if it came to that.

Tristan Harris: I think we need to get good at holding each other through to the post-tragic. I don't know what that looks like, but I know that that's what guides me and what we're trying to do. And if there's anything that I think I want to get even better at is, it's hard once you take people through all these things to carry them through to the other side.

Kara Swisher: Right. Because they get hopeless. They get hopeless. Yeah, you can be hopeless. After that thing, I came back, I'm like, "We are fucked." And I thought, that's not going to go over well because most people hide in on Instagram or TikTok.

Tristan Harris: That doesn't feel good. Let me run away from myself again. Let me scroll a bunch of photos.

Kara Swisher: Exactly.

Tristan Harris: This is going to be a difficult time. The more we can go through and see the thing together; I think part of being post-tragic is actually going through it with each other, being there with each other as we go through it. I'm not saying that just as a bullshit throwaway line, I really mean it. I think we need to be there for each other.

Kara Swisher: All right. Post-tragic, hand in hand. Here we go.

Tristan Harris: Post-tragic, hand in hand. Let's do it.

Kara Swisher: Tristan, thank you.

Tristan Harris: Okay, thanks.

Kara Swisher: Bye.

Speaker 3: Today's show was produced by Nayeema Raza, Blakeney Schick, Cristian Castro Rossel, and Megan Burney. Special thanks to Mary Mathis. Our engineers are Fernando Arruda and Rick Kwan. Our theme music is by Tracademics.

If you're already following this show, welcome to the world of post-tragedy. Hey, it could be worse. If not, it's a high P(doom) for you. Go wherever you listen to podcasts, search for *On with Kara Swisher* and hit follow. Thanks for listening to *On with Kara Swisher* from New York Magazine, the Vox Media Podcast Network and us. We'll be back on Monday with more.