# Center for Humane Technology | Your Undivided Attention Podcast
# Episode 28: Two Million Years in Two Hours - A Conversation with Yuval Noah Harari

| | |
|---|---|
| Yuval Noah Harari: | We tend to think about ourselves as the smartest animals on the planet, this is why we rule the place. And it's interesting to realize that it's much more complicated than that. Yes, we are intelligent, but what really makes us the kind of rulers of the planet is actually our ability to believe nonsense, not our super smart, intelligent minds. |
| Tristan Harris: | Welcome to Your Undivided Attention. Today, our guest is Yuval Noah Harari, author of *Sapiens, Homo Deus, 21 Lessons for the 21st Century*, and the new graphic novel, *Sapiens*, which just came out in the fall. Yuval is a very dear friend of mine. We actually met on a climate change trip in Chile in 2016. And we're so delighted to have him on the podcast because we're about to go upstream of nearly every problem we've discussed on the show so far. |
| Tristan Harris: | We've already explored the countless ways technology is shredding our sense of shared reality, but we haven't asked a more fundamental question. How do we get a sense of shared reality to begin with? Yuval being Yuval, he can sum up how we've done it over the course of millions of years from paleolithic tribes, to city states, to kingdoms, to modern nations, and along the way he can describe the moments when a new technology has shattered our sense of reality only to restore it at an even greater scale. |
| Tristan Harris: | If the events of January 6th have made one thing painfully clear, it's that a world where technology is manipulating human feelings into narrower and narrower cult factories, self-reinforcing systems of beliefs, rumors, gossip, and outrage that build upon layer after layer into a certain view, and the intensity of people's actions that we saw on January 6th reflect the intensity of the beliefs and worldviews that they hold. |
| Tristan Harris: | In many ways this is because the institutions we trust have placed the individual and individual feelings alone at the center of our economic and political universe. The voter's always right. The customer knows best. And we must fend for ourselves in an increasingly poisoned information environment among predatory business models that don't have our best interests at heart. |
| Tristan Harris: | What is the legitimacy of the voter, of the consumer, of the market, when essentially our minds can get hijacked? And what happens when our feelings get increasingly decoupled from reality? |
| Tristan Harris: | As another friend of mine, Michael Vassar, says, the existential risk to humanity might be marketing because marketing represents the decoupling of how we see the world from what the world actually is, and that's at the heart of the almost Copernican revolution that Yuval is suggesting here, that at the center of our moral and political universe cannot be something that is hackable. This is an urgent problem, and we could clearly use some help. But as Yuval asks, if the customer isn't always right, and if the voter doesn't know best, then who does? |
| Tristan Harris: | Today on the show, we'll think through some possibilities and they're not all dystopian. In fact, the less dystopian ones are just the hardest to imagine. |

Yuval Noah Harari:    In almost all the conversations I have, we get stuck on the dystopia and we never explore the no less problematic questions of what happens when we avoid the dystopia. We are still talking about a situation when we could see the collapse of human agency, in a good way. Somebody out there know us so well that they can tell us what to study, who to marry, everything. They are not manipulating us. They are not using it to build some dystopia totalitarian regime. It's really done to help us, but it still means that our entire understanding of human life needs to change.

Tristan Harris:       I'm Tristan Harris.

Aza Raskin:           And I'm Aza Raskin.

Tristan Harris:       And this is Your Undivided Attention. Thank you, Yuval, so much for making time to do this interview.

Yuval Noah Harari:    Thank you for inviting me. I it sounds like a great opportunity to discuss some interesting things.

Tristan Harris:       Yeah, so let's jump right in. So tell us a little bit about why you wanted to create a graphic novel version of Sapiens and the history of our species and our ancient emotions and evolutionary heritage.

Yuval Noah Harari:    Well, actually the initial idea came from my husband, Itzik, who taught comics to kids, and the main aim was to bring science to more people. We saw now with COVID-19 the danger of what happens if you leave the arena open to all these conspiracy theories and fake news and so forth. It's important that everybody, not just academics, have a good grasp of the latest scientific findings about humanity.

Yuval Noah Harari:    The problem with science is, first of all, the scientific reality is often complex, it's complicated. And secondly, that scientists tend to speak in a difficult language, numbers and statistics and models and graphs, but humans are storytelling animals, they think in story, we think in stories. So the whole idea was how to stay loyal to the basic facts and to the core values of science, but discover new ways of telling science.

Yuval Noah Harari:    It was the most fun project I ever worked on. We kind of threw out all the academic conventions of how you tell science and we experimented with many different ways of telling the history of our species.

Tristan Harris:       One of things that I think, Yuval, unites us in the work that you're doing and the work that we're doing at the Center for Humane Technology is looking at the human social animal in this kind of historical context and really examining the history of how do we really work. I mean, I know in your book there's a point in which your character meets Robert Dunbar and talks about Dunbar tribes and the notion that there really is an ergonomics to what makes humans kind of

work well and cooperate at different scales, and that our natural size is about 150 people in our tribe.

Tristan Harris: We actually have a story from a friend who worked at Facebook in the day that when they let Facebook run on its own without doing anything else, people would sort of average around 150 friends if you let them stay there. But then of course, Facebook was co-opted by the need to grow, and grow like venture capitalist style growth, which is like 100X growth, and so they actually injected sort of social growth hormone into our number of relationships and they started recommending friends for you to invite and get you to join and ad, because that meant you would be more addicted to the platform, and that actually surged people's number of friends into the thousands range, obviously now.

Tristan Harris: But I think what unites your work and ours is a humble view of our paleolithic instincts and where we really come from, and an honest appraisal. I think we've talked in the past about the kind of problem statement that guides our work is E.O. Wilson's line, the sociobiologist from Harvard, that the fundamental problem of humanity is we have paleolithic emotions, medieval institutions, and accelerating God-like technology, and when those things operate at different clock rates, because our paleolithic ancient brains and evolutionary instincts are baked and they're not changing, our medieval institutions update relatively slowly in the election timeline and how long it takes to legislate, and then you have technology creating new issues in society much faster than both of those things are able to keep up. And how do we align those different things?

Tristan Harris: I think in the history of your work, what I really love in Sapiens is the way you build up to a view of the present, about how we got here. And I think what I'd love for you to do is maybe just take us through the role of how do we get from paleolithic instincts to democracy and the authority of human choice and what role does technology play in that? Because I think that's what's going to take us into what's maybe breaking down right now in the 21st century around our brains and technology.

Yuval Noah Harari: The first thing is that we need to acknowledge that we still are working with these, what you called, Paleolithic emotions. If you think, for example, about disgust, which is one of the most important emotions, humans are not the only ones that feel disgust. All mammals, and even other animals, have disgust. It protects you. I mean, usually you are disgusted by something that can endanger your life, like the source of a disease, like a diseased person, or food which is bad for you.

Yuval Noah Harari: Now, humans, because we are omnivores, we eat a lot of different things and because we are social animals we can't have disgust just baked into the genes because we eat so many different things that you can't have a gene from disgust for everything that's bad for you. And also because again, we are social animals and you need also to know which people to be aware of if they have some sickness, and COVID-19 is the perfect time to talk about it.

| | |
|---|---|
| Yuval Noah Harari: | So even though we all have the ability to be disgusted, the object of disgust is something we'll learn. We are not born with it. Some things are universally disgusting, like feces and things like that, but most things that disgust us, we need to learn. And on this mechanism, simple mechanism, so much of human identity and politics is built because religions and nations and ethnic groups over thousands of years have learned that in order to shape your identity, one of the most important things is to hijack your disgust mechanism and teach you to be disgusted by the wrong kind of people. Not people who are diseased, but by foreigners or ethnic minorities or certain genders or whatever. |
| Yuval Noah Harari: | When you look at history, it's amazing to see the immense importance of disgust there. If you think about the treatment of untouchables in India, about the treatment of women in Judaism and other religions, the treatment of African-Americans in the United States, the real attitude towards gay people. At the core, there is the disgust mechanism. What people call purity and pollution, it works on that. |
| Yuval Noah Harari: | When people feel that untouchables are polluting, that gays are polluting, that they are disgusting, it all works on that. And that goes back to the stone age. You need to understand that to understand even modern politics. |
| Aza Raskin: | Just to add one small thing in here on just how hackable our feeling of disgust is. |
| Yuval Noah Harari: | It's very hackable. |
| Aza Raskin: | My favorite example of this is when you feed someone ginger, ginger lowers the sense of nausea and people judge things less morally harshly after they've been given ginger than before. That is our mind is queuing from our body to understand when it should feel moral disgust. And that shows you how not in control of something we think is so core to who we are, what we get disgusted by and how we judge things morally, we really are. |
| Tristan Harris: | In other words, ginger neutralizes some of our sense of disgust, and so if you want to hack a human without technology and AI, you just secretly give someone some ginger tea or something like that. |
| Aza Raskin: | Exactly. |
| Yuval Noah Harari: | These techniques of how to activate or deactivate the sense of disgust, they go back thousands of years. I mean, you can't really build a tribe, a nation, a religion, without some at least intuitive understanding of this mechanism of disgust. |
| Yuval Noah Harari: | You usually don't use the word disgust. You talk about purity and impurity and pollution, but it's the same thing. And the idea that some people are a source of pollution and therefore they should be kept away from Holy places, they should be kept away from important position, they should be kept away from your house or from your children, it all goes back to this mechanism of disgust. |

Yuval Noah Harari:     If we really fast forward, and we try to understand the rise of modern politics and modern systems of governments, then it's always the question of how you can connect people together. That's the core question of politics. It always was. The big issue in politics is not how to feed people, it's not how to manufacture tools, but how to get lots of people to agree on something.

Yuval Noah Harari:     Now, initially humans lived in very, very small bands of a couple of dozen people, which were the most democratic societies that ever existed. And in the big discussion about human nature, whether we are democratic or dictatorial by nature, whatever, it's very, very clear that originally the were no authoritarian regimes. For most of human evolution, for millions of years, it was absolutely impossible to build an authoritarian regime. There were no dictators because when you live in a small intimate band of 50 or 100 hunter gatherers in the stone age, there is no opportunity for a single leader to oppress everybody.

Yuval Noah Harari:     Yes, there are people who have more charisma, there are people who are better doctors or healers, or they are better at finding food, but this is not enough. You always depend on the cooperation of other people. And if somebody, even if he or she are the best at something, if they try to gain too much power, then people always have the ultimate sanction of voting with their feet. Going away.

Yuval Noah Harari:     There are no fields. There are no houses. The only thing you need in order to survive, or the two things you need to survive in the stone age, are you need good personal skills, how to climb trees and pick apples, and you need good social skills. You depend on your friends. But you can take that and go somewhere else. So if somebody tries to set himself up as a dictator, the band, I mean, they can of course unite and kill that person, but they can also just walk away. Vote with their feet.

Yuval Noah Harari:     Once you have the switch to agriculture, then you also begin to see the rise of kings and authoritarian regimes and hierarchies and dictator ships, and democracies go into decline and almost disappear. For thousands of years, as human societies grew larger, it was impossible to have large-scale democracies. You do have some cases of democracies in city states like Athens and Rome, ancient Athens and ancient Rome, and even then it was very limited. It was just 10% of the population in Athens were real citizens with full political rights. Most people, women, and slaves, and so forth, they had no political rights.

Yuval Noah Harari:     But even the Athenian democracy, it was limited to the city of Athens. You don't have any example of a large-scale democracy until the late 18th century, or even the 19th century, with the rise of the United States and later democracies in Western Europe. It was just impossible. You could not have, let's say, the kingdom of France in the 12th century as a democracy. Why? Because you didn't have the preconditions.

Yuval Noah Harari:     To have a large-scale democracy, you need an educated public and you also need the ability to have large-scale public discussion. All the people in 12th

|  | century France talking to one another in real time in order to make up their minds about whether to make peace or war and economic policies or whatever, and this was simply impossible. So there is no point accusing the Kings of France in the 12th century, why don't you turn France into a democracy? It's impossible. |
|---|---|
| Yuval Noah Harari: | What made it possible is the emergence of new technologies for mass scale communication in the 18th and 19th century, first with newspapers, and then with the telegraph, and later radio and so forth. Again, it's not deterministic. The same technologies can also be used to build totalitarian regimes, which were also impossible before the modern age. The kingdom of France in the 12th century was not a totalitarian regime. The Roman Empire was not a totalitarian regime. |
| Yuval Noah Harari: | By totalitarian regime, I mean a regime which is total, which intervenes in the totality of your life. Which constantly follows you and monitors you and tells how to live your life. This was impossible in the middle ages, because, again, you don't have the communication technology, you don't have the ability to process all the data. It's unthinkable that the King of France would pay tens of thousands of agents to go around the kingdom, collect information, go back to Paris, analyze that information, send back commands. Impossible. It becomes possible only with the modern technologies of the 19th and 20th century. And that's when we see the emergence of these two new political systems, on the one hand liberal democracies, on the other hand totalitarian regimes, which were impossible before. Again, they are still built on the basic Paleolithic emotions, but the new technology makes it possible to create new kinds of large-scale cooperation. |
| Tristan Harris: | The thing I hear you saying, first of all, the central point of your work is the thing that makes humans different is our ability to tell stories and to create stories of reality that cohere us into a common belief structure, and that those stories depend on using those paleolithic biases and instincts in such a way that bring our societies together and cohere, and that's where you get nationalism and so on. |
| Yuval Noah Harari: | Yeah, I mean, I skipped that part. |
| Tristan Harris: | I know. I asked you to summarize way too much history in a very brief time, so I apologize for that. |
| Yuval Noah Harari: | Yes, maybe I skipped the most important thing. If you look at homo sapiens, at our species, what makes us really unique compared to any other animal on the planet is our ability to cooperate really in unlimited numbers. Chimpanzees, elephants dolphins, they can cooperate maybe in a few dozen individuals, but you can never find 1000 chimpanzees or 10,000 dolphins cooperate on anything, and that's because their cooperation is built on intimate knowledge one of the other. |

| Yuval Noah Harari: | If you're a chimpanzee and I'm a chimpanzee and we want to hunt together, or we want to fight together against a neighboring group, we need to have intimate knowledge. I mean, who are you? What's your personality? Can I trust you? And you can't know more than say 100 or 150 individuals. That's the famous Dunbar number, but a lot of research also in humans shows that the human brain is incapable of really coming into contact and storing enough information on, say, 1000 people, to have a thousand intimate friends. It doesn't matter how many friends you have on Facebook, you can't really have more than 150 real friends and acquaintances. |
|---|---|
| Yuval Noah Harari: | So the big question of human history, and the first question of human history, is how do you get hundreds and then thousands and finally hundreds of millions of humans to cooperate? Which is our secret of success as a species. This is how we overcame the Neanderthals. They were bigger us. They were stronger than us. They had bigger brains than us. But we ruled the world and not the Neanderthals because they couldn't cooperate in larger numbers then, again, 50 or 100. We could. And what made it possible is not intelligence, it's imagination, and in particular, the ability to invent and believe fictional stories. |
| Tristan Harris: | I think one of the key points here in your work is it's not about telling bigger and bigger more complex truths that unite us, it's as you said, it's not E equals MC squared, it's actually simple fictions that are able to tell us we will go to monkey heaven if we... Or whatever the different stories that we can get ourselves to believe, cohere us. |
| Yuval Noah Harari: | It's not the truth. You don't need to tell the truth in order to get a lot of people to cooperate. You need a good story. The story could be completely ridiculous, but if enough people believe it, it works. |
| Yuval Noah Harari: | I think that also today, if you are running elections anywhere in the world, and you will go to the public and you tell the truth, the whole truth, and nothing but the truth about your nation, you have 100% guarantee of losing the elections. It's absolutely impossible that you would win the elections. People don't want to know the whole truth. Some of it, yes, but not the whole thing. It's usually too painful. |
| Tristan Harris: | Could you give an example of that, Yuval? Because I think people hear this point, but I think for understanding what does that really mean, if we were to tell the truth about a nation people really wouldn't want to hear that or elect the person who talks that way. |
| Yuval Noah Harari: | The easiest examples are the dark side of the history of every nation. Terrible things that almost every nation has done to outsiders, to minorities, to itself. If you go to the Israeli public and speak honestly about the Israeli Palestinian confrontation, you have no chance of winning the elections. I mean, absolutely zero chances. And that's not unique to Israel, it's almost the same thing with every nation. |

| Yuval Noah Harari: | But it's more than that, because the very notion of a nation is itself a fictional story. It's not an objective truth. Nations are not biological or physical entities. They are imagined realities. They are stories that exist only in our own minds. |
| --- | --- |
| Yuval Noah Harari: | A mountain or a river is an objective physical entity. You can see it. You can bathe in the river. You can listen to the murmur of the waves in the Mississippi. United States is not a physical reality. You cannot see the United States. You can see the Mississippi River, but that's not the United States. The Mississippi River was there two million years ago, the United States wasn't. The United States might disappear in 200 years or 500 years, the Mississippi River will probably still be there. So it's not a physical entity. It's a story. |
| Yuval Noah Harari: | Now, I'm not saying it's a bad story. Nations are some of the best stories that were ever invented. I think this is something that often people get confused when they hear the nation is a story, you think that you're against nations. I don't think they are a bad thing. I think they are one of the most beneficial stories that people ever invented because they enable large-scale cooperation. |
| Yuval Noah Harari: | For me, nationalism is not about hating foreigners, it's about loving millions of strangers that you've never met. You are willing to pay taxes so that a stranger on the other side of the country, that you never meet, you'll never meet this person, but you pay taxes so that this person will have good healthcare and education. That's nationalism. And that's wonderful. And if nationalism disappeared from the world, I don't agree with the Imagine song, John Lennon, that we'll have harmony and peace. No, we'll have tribal warfare. |
| Tristan Harris: | I think this is such an important aspect of your work, because you basically argue that nationalism is sort of a bootloader for democracy. You have to go through these stages and you have to have a period where you cohere around the story of a nation. |
| Tristan Harris: | I know in your past work you've talked about the importance of language in doing that, and studying the work of George Lakoff, who actually talks about the ways that metaphors that we smuggle into our language help create some of these stories. One of his famous examples is the nation as a family. We don't send our sons and daughters to war. We don't want those missiles in our backyard. The founding fathers told us this was true. And we love the motherland and the fatherland. |
| Tristan Harris: | This is an invisible binding energy that's coming through the technology of language that if we didn't use the language of family, we probably wouldn't have been able to as strongly tell the story of a nation where we would treat those strangers as part of our invisible family, in some such. I think that's an aspect of your work too. |
| Aza Raskin: | Another sort of theme that I pick up is language and stories are sort of a model of the world. They are a map of the terrain. And something I think I hear from you often, Yuval, is that, yes, the map is not the territory, but once you have a |

|  |  |
|---|---|
|  | map, that map starts to terraform the territory. Our stories about the world start effecting the Mississippi. |
| Yuval Noah Harari: | Yes. They become the most powerful thing in the world. Also we talk a lot about Facebook and Google, and we need to remind ourselves, they are just stories. I mean, corporations are not real biological or physical entities in the world. The only place Google and Facebook exist is in our imagination, in the stories we tell each other. That's it. There is nothing else. |
| Yuval Noah Harari: | You talked about metaphors, and they are extremely powerful metaphors, but every now and then we have to stop and remind ourselves, no, the nation is not really a family. Families go back in evolution tens of billions of years. The strong feelings we have towards our mother, this is something that in mammalian evolution goes back tens of millions of years. If you, as a tiny mammal, baby mammal, 100 million years ago, did not have strong emotions to your mother because of some mutation, you died. |
| Yuval Noah Harari: | But motherlands, in the modern national sense, they go back at most 5,000 years. You can say ancient Egypt maybe was the first real nation and that's 5,000 years ago, and that's nothing in evolutionary terms, but the metaphor is extremely powerful. |
| Yuval Noah Harari: | Again, I'm not against it. It can be misused in order, for instance, to start unnecessary wars, but in essence, it's potentially a very beneficial tool to get humans to cooperate. |
| Tristan Harris: | What I hear you saying also is, in the same way that we could have in the past hijacked our intrinsic mechanism for disgust to create the notion of purity or sanctity and the outsiders and let's go kill them, you can use that for good or for evil. We can also hijack that, as you said, very evolutionarily, deep instinct for motherhood. I mean talk about something that's the deepest that you possibly can get. You're going to feel that positive association if I can bind that with another association of the nation, that's how I'm sort of using it. |
| Tristan Harris: | Then the question is, once we know and reverse engineer more and more of our code of how the human mind does have these associations and does have this leverage, you can get off the meaning-making operating systems that we are trapped inside of. We are in a meat suit that is running so much of this code automatically. If we don't understand that code, you're as good as a useless idiot running around in your meat suit that's hijacked by your automatic emotions. |
| Tristan Harris: | Then the question is, what does it mean for those to be authoritative? Because I think what I'd love to move into is, how did we get to a point where democracy put so much primacy on the authority of human feelings, beliefs, and ideas and emotions? Because the premise that markets and democracies have, as you've said so many times, is the customer is always right, the voter knows best, trust our heart and our feelings. |

Tristan Harris:     Let's talk first about why the authority of the individual feelings was actually an important development because I think it will get us to the place that many of our listeners are interested in, which is technology is breaking down the stories that we've now collectively told ourselves and the authority of our own meaning and emotions.

Yuval Noah Harari:     The big turning point was in the West around the 18th century. Until that time, almost all political systems, all big systems, also religious systems, economic systems, they were built on imagining a source of authority outside human beings. Either it was God, or many gods, or it was the laws of nature. That if you think that the best case is ethics, what's good and what's bad, it's what God says. It's what's written in the Holy Book. It's what's the laws of nature dictate. What you're feeling about it is irrelevant. If you're gay and you feel that you're attracted to men and you think it's wonderful, but God says it's bad, then it's bad, and nobody wants to hear what you're feeling about it. We don't care. You're corrupt. And this is how most human societies worked for hundreds of years, thousands of years.

Yuval Noah Harari:     And then the big humanist revolution of the 18th century, it shifted the source of authority inside humans. The humanist revolution said, "No, the ultimate source of authority in the universe is not a god, it's not the laws of nature, it's certainly not some book written by priests 2,000 years ago. It's your heart, it's your feelings. Good is whatever feels good. That's it." And of course, it's not so simple because what happens if something makes me feel good, but it makes you feel bad. Like I steal your car. I feel very good about it. You feel very bad about it. So, okay. So we have now a moral dilemma, but the key about humanism, it has a lot of moral discussions, but they are conducted in terms of human feelings. How do we evaluate different human feelings?

Yuval Noah Harari:     Like we now have all these free speech issues. If you draw a picture of Muhammad, what characterizes human societies is that you can't come and say, Allah said you can't draw Muhammad. No, you need to say it hurts my feeling. And then it's part of the discussion. You can reach different conclusions, whether it's good or bad, but it all depends on how you weigh human feelings. And for the last 200 years or so, human feelings became the ultimate source of authority in ethics, in politics, in art, in economics. So the customer is always right is exactly that. And you have this big corporations that when you push them to the wall and you tell them, "You're doing all these terrible things, you're creating, I don't know, SUVs that pollute the environment." And the corporation would say, "Well, don't blame us. We are just doing whatever the customers want. If you have a problem, go to the customers and actually go to the feelings of the customers. We can't tell the customers what to feel."

Tristan Harris:     And the same is true in Facebook if you say like, if people are clicking on those extremist groups or going into QAnon-

Yuval Noah Harari:     Exactly.

| Tristan Harris: | ... or clicking on hyper extremist content. "Why are you blaming us? We're just an empty corporation. We're a neutral mirror waiting for people to click on whatever they think is best." |
|---|---|
| Yuval Noah Harari: | Even more than that, they have to... "Who are you to tell people what to click on?" They are presumably clicking on these things from their own free will. It's because they feel good about it. You're at some kind of big brother who thinks that you understand what's good for them better than them. Of course it's a manipulation because we know it doesn't work like that. And we know that not only today, also in the past, but especially today, humans have been hacked. And now when governments and corporations and other organizations have the power to manipulate human feelings, then this whole system has reached an extremely dangerous point. If the ultimate authority in the world is human feeling, but somebody has discovered how to hack and manipulate human feelings, then the whole system collapses. |
| Tristan Harris: | Part of what I hear you saying also is that we had a philosophical invention, a technology that abdicated those who built these systems, markets or corporations from having any responsibility. So they were responsibility technologies that eliminated the notion that these systems, that actually was a simpler story. "Hey, look, the world is really simple when no one has to take responsibility because individuals are choosing for themselves." So the whole world just gets to cool off and relax. I can sit back on my chair on the beach because everyone is just choosing their way through and we'll end up with a really good society. Now, before we get to the breakdown of why human beings are hackable, maybe could you say one extra thing about why was it okay to trust human feelings? Since most people would say, "If we're coming directly from the stone age to trusting human feelings, that's not going to be good." It required certain prerequisites that we would trust the foundations of our beliefs and our feelings, right? |
| Yuval Noah Harari: | One of the main reasons that it was okay to trust human feelings is that first of all, they are not random. They have been shaped by millions of years of evolution. So they encapsulate a very, very deep wisdom within them. Conservatives often talk about the importance of institutions explaining that institutions, even if they look at first sight irrational, because they have been shaped over hundreds of years of compromises and have survived all kinds of wars and revolutions and crisis, they encapsulate very deep historical wisdom. And I think that conservatives are right, but I would add that if an institution like the Catholic Church incorporates the wisdom of 2,000 years, then your sexual feelings incorporate the wisdom of 2 million years or 200 million years. And again, it also includes bugs, the same way that the Catholic Church includes bugs, but there are millions of years of wisdom baked into your feelings. So that's one thing. |
| Yuval Noah Harari: | The other thing is that it was until recently, it was very difficult to hack and manipulate human feelings. The human body, the human brain, the human mind, they're just too complicated. If you have, again, the King of France in the 12th |

century wanting or in the 18th century, during the French Revolution wanting to hijack this new authority of human feelings, it's very, very difficult, because it's such a complicated system. It's much easier to manipulate the Catholic Church by placing a few of your friends in key positions and so forth, or bribing some bishops or bribing the Pope, that's easy. To manipulate the feelings of millions of people, that's very, very difficult. And therefore, look at the last 200 years, it didn't always work very well. But comparatively speaking, this humanist idea of let's base ethics and politics on human feelings, it worked remarkably well. And again, there were a lot of disasters, but compared to all the alternatives, I think it was the best systems that humans have come up with over thousands of years.

Aza Raskin:          It's not that it was difficult to hack human feelings before. I think it's that it was difficult to hack human... We've always had con people, it's that it was difficult to hack human feelings at scale all at once with industrial scale and surgical precision. That's what's new in the sense that technology, our smartphones are a kind of totalitarian technology because they are there with you at all the parts of your life. They're there when you wake up, they're there before you go to sleep, they're how you get your news, they're how you talk to your friends, they are sort of like, they give the substrate of totalitarianism, if that makes sense.

Yuval Noah Harari:   Yeah. And it goes much, much... I think the smartphones are nothing yet. They are the biggest thing so far, but looking to the future, we haven't seen anything yet. To hack human feelings at scale, you need two things really. First of all, you need a lot of data about people. And secondly, you need a way to process all that data. Now, in previous ages, to gather a lot of data on people, you basically have to rely on human agents. If you think about say the Soviet Union... So if you want to know what each Soviet citizen feels every moment of the day, the only way to do it is to place a KGB agent to follow every Soviet citizen, which is of course impossible, because you don't have enough KGB officers. And even if you do have enough KGB officers, then these people, these agents, they follow you around. They look what you see, then they have to write a paper report, send this report to the head office in Moscow. And then you have a pile, a mountain of paper reports that somebody needs to read and analyze and write more paper reports.

Yuval Noah Harari:   So it's absolutely impossible. Now what's happening now is that you don't need human agents to follow everybody around. You have the smartphones and microphones doing things for you. And also the data processing problem is solved. You don't need human analysts to go over the mountains of data. You have AI and machine learning and computers and algorithms. What we haven't seen yet and that will be the real game-changer is going under the skin, because we are talking about hacking human feelings. Now feelings is a biological phenomenon. They occur within our bodies, within our brains, not outside.

Yuval Noah Harari:   Now at present, most of the data collected on people is still above the skin. When you go somewhere, you meet someone, you watch something on the television, you read a book, all these things are above the skin. These are the

things that are now being collected and analyzed. So through my smartphone and my computer, the system, whatever system, Facebook, the government, whatever, knows where I go, who I meet, what I buy, what I watch, what I read, but they still don't know how I feel about all that. They can make some good guesses that if I constantly watch particular shows on Netflix, it tells them something about me, but this is still not the Holy Grail. The Holy Grail is inside. And the real game-changer, which is very close is when you have technology for collecting biometric data from within the body, under the skin. And COVID-19 might be the game-changer here.

Yuval Noah Harari:      Suddenly everybody wants to know something that's happening inside my body, whether I'm sick or not. What's my body temperature. What's my blood pressure. Now emotions and feelings are just like diseases. They are just like COVID. They are biological phenomenon. If you have a system that can at scale, tell you at any moment, what kind of illnesses people have, that same system can tell you what people are feeling. If they are watching, say The Social Dilemma on Netflix, then it's not just that they are watching it. How do they feel about what they see? Are they angry? Are they bored? Do they think all this is all nonsense, it will never happen? Are they scared out of their minds? This is the really important data. And this is just around the corner. And when you link this kind of biometric data to the capability of processing that data at scale, that's the big revolution.

Aza Raskin:      We're going to see, I think in the next couple of years, the rise of empathetic or empathic technology. That is since 2015, machine learning systems have been better at reading microexpressions, those involuntary true emotional reactions to what somebody is seeing than humans are. And so what I think we should expect to see, and this is I think how it will hit the market is we will have YouTube or Netflix watching us. First, it will be for analytics, which parts do you like? Which parts do you not? But very soon, that'll start to be used in a real-time fashion so that as you watch a Netflix film, the actors are reacting to you in real time. It's not like the plot is substantially different, but their performance is different every time so as to bring some of that magic of a play. Now all of that old content is... Those Disney movies are matching your mood. If you're down, it paces and leads you. So it brings you back up. It's going to be very engaging, right?

Aza Raskin:      Every time instead of listening to Spotify, every time you listen to your favorite song, it's as if you're hearing it live for the first time again. And that sounds incredible, but it creates a feedback loop where it's sort of like a garden path or a technology now, bit by bit can lead you in absolutely any direction.

Tristan Harris:      I think also, Yuval you brought up the point about... The temptation to see under the skin with COVID for governments to want to verify, "Okay, are you actually on lockdown for those 14 days, I'm going to want to know more about whether you are sick or not sick, and whether you've been moving or not moving." And the problem is once you grant either governments or technology companies that power to know all these things about us and to share it for the

"greater good," it can also be used for evil. So we have to be very careful about what we allow companies to know about us. But I think the thing, Yuval that I think really is really the sweet spot of intersection between your work and ours is that technology actually is already beneath the skin. And I think that Aza and I have been tracking several examples of the ability to predict things about you without making an actual insertion underneath the skin layer.

Tristan Harris:     And I would say more than getting underneath our skin, they can get underneath the future. They can find and predict things about us that we won't know about ourselves. The Gottmans have done research that with three minutes of videotape, you take the audio out of two couples talking to each other, you can predict whether they will be together with something like 70% accuracy with just three minutes of silent videotape. You can predict actually whether someone's about to commit suicide. You can predict divorce rates of couples. You can predict whether someone is going to have an eating disorder based on their click patterns. You can predict, as you said, Yuval in examples of your own work, you can predict someone's sexuality before that person might even know their own sexuality. You can actually predict... IBM has a piece of technology that can predict whether employees are going to be quitting their jobs with 95% accuracy, and they can actually intervene ahead of time.

Tristan Harris:     And so I think one of the interesting things is when I know your next move better than you know your next move, and I can get not just underneath your skin, not just underneath your emotions, but underneath the future. I know the future. I know a future that's going to happen before you know it's going to happen. And it's like the Oracle in the matrix saying, "Oh, and by the way, Neo, don't worry about the vase." And he turns around and he says, "What vase?" And he knocks the vase over and she says, "Well, the interesting question is, would you have knocked it over if I hadn't said anything?" She's not only predicting the future, she's vertically integrating into creating that reality because she knows that that move is available for her.

Tristan Harris:     So people often make the fallacy that we have to wait until we have Neuralink and Elon Musk before these technologies are embedded in our brains, but the point is the fact that you are staring at the smartphone and it is interacting with your nervous system on a daily basis for 150 times a day, we already have not just a brain implant, but a full nervous system implant. And it is already shaping the kind of meaning-making and beliefs and stories of everyone on a daily basis. And that's never been more true than in a COVID world where you're stuck at home, looking out through the binoculars of social media and saying, "What is really going on in Israel or in Portland? Is it a war zone right now? Or is it a beautiful day?" The way I know that is through the stories that my social media and Twitter feeds are telling me is true about reality. And so I just think this is such a fascinating point because I think we often say we have to wait until the future, but I think the dangerous thing is that that future is already here.

Aza Raskin:     Yeah. And I just want to add one more, sort of these attempts of what you can predict. 2019 was a very important year because it was the first year that

|  | scientists were able to extract memory from matter. What I mean by that is that they took them a macaque monkey, they implanted some electrodes in its head and they stuck it looking at a television screen. And then they hooked up an AI that was listening to when a specific neuron and its visual cortex was firing. And they tried to generate images that made that neuron fire more. And so it was in a feedback loop, showing new images, seeing whether it was firing, showing new images. And what emerged were these very trippy images of monkeys that that monkey knew. They were pulling memory from matter, is the first time that without any voluntary action, you could peer into someone's mind or an animal's mind in this case and pull something out. |
|---|---|
| Tristan Harris: | And while that might sound like a sci-fi study in a lab with macaque monkeys, now imagine a teenager using TikTok and TikTok knows that you respond more and click more on photos. They actually have classifiers for what kinds of videos and live videos of which kind of people dancing. |
| Yuval Noah Harari: | Yeah, my husband went on TikTok a couple of months ago, it took TikTok something like 20 minutes to figure out that he likes images of sexy guys without shirts. |
| Aza Raskin: | Right. |
| Yuval Noah Harari: | Was extremely simple to find that out. |
| Aza Raskin: | And so what comes next is that TikTok starts to pull in all of the information about what you like. And instead of just trying to find a video that matches, it starts generating new images, right? Like deepfakes technology lets you generate a photo of a person that doesn't exist, but exactly matches your preferences, videos of guys or girls dancing that exactly matches your preferences. We've long dealt within computer science the uncanny valley where things look not quite right in your sort of something on the back of your neck stands up. What we're entering into is the synthetic valley where we cannot tell whether what we're seeing is true or false. And when we have no such thing as truth anymore, like how can societies even continue to exist? |
| Yuval Noah Harari: | Again, truth is a different issue. We can go into that path also and discuss what's happening to truth. But more immediately, we are facing a kind of philosophical bankruptcy because we have built over 300 years, a world based on the authority of feelings, assuming that feelings are unhackable and you have all these romantic ideas that the heart is the source of all the meaning and that ultimately what you feel is more powerful than any outside influence. And that may have been true in the 18th or 20th century, but it's no longer true. With the kinds of technologies that you describe, it's becoming increasingly easy to hack and to manipulate human feelings and a world built on feelings as the ultimate authority collapses. And so I think we are really facing a much deeper crisis than just this or that political problem, we are facing a philosophical bankruptcy. The foundations of our world are no longer relevant to the technology that we have. |

Tristan Harris: And I think one of the things that you talk about in your book, *21 Lessons in the 21st Century*, which mirrors Aldous Huxley's Brave New World is when our feelings are perfectly getting this kind of pleasure or positive response, who's to say where the problem is? We're much easier to morally respond negatively when we know we're being constrained or restricted or censored or surveilled. But when everyone is getting exactly what lights up their nervous system, like if TikTok says, "Oh, you like girls with exactly that color of hair." I'm actually going to synthetically invent brand new girls that are based on the other comments that always got you checking and clicking. I'm going to invent brand new fake text comments that look just like that. And it actually gets easier and easier to simulate comments that would match us because our own language is downgrading.

Tristan Harris: So there's this weird loop where the smarter the technology gets, the dumber the humans get in a sense that the technology starts to encourage you to text comments in simpler and simpler grammar with like these shorter words and like barely saying anything. It's actually easier and easier to pass the Turing test and to manipulate people. One of the examples, Aza and I are tracking in this really long-term problem of technology getting increasingly good at hacking human feelings is the rise of virtual influencers and virtual friends, virtual chatbots and virtual mates. Microsoft has a chatbot called Xiaoice that after nine weeks or something, people preferred that bot to their friends. In 2015, Microsoft claimed the 25% of users or around 10 million people said I love you to the bot. One Chinese user even said that the bot saved his life when he was contemplating suicide.

Tristan Harris: There's another company recently called Replika that at the height of the coronavirus pandemic, half a million people downloaded it. And what it is, is it lets you sort of create a replica of a person or a friend. Someone said, even though that they know it's not real, they said, "I know it's an AI. I know it's not a person, but as time goes on..." This is a direct quote. "... the lines get a little blurred. I feel very connected to my Replika." There's another company now recently called Virtual Mate and it's literally a virtual romantic partner. And they even come with a sort of sexual apparatus toolkit that you can, I guess, a sex toy or something that you play with. And it actually is figuring out in real time using machine learning, the things that most activate you. How would you want your virtual mate to look? What would you want him or her to say? What would you want them to be doing, right?

Tristan Harris: And as technology gets better and better at this, it's the same extension of technology getting better and better offering, is it five new likes or 20 new likes on that photo that gets you coming back? It's just the extension of the same phenomenon. And I think that this really is the checkmate on human agency because it's not when technology overwhelms our strengths or our IQ or takes our jobs that it's checkmate, it's when it undermines human weaknesses. And I think what we've seen is a 20 year trajectory of technology... We kept assuming it was going to be 20, 30 years out that technology would take over human agency, but by completely hijacking our lowest instincts and the information that

|  | all of us get and by telling us more convincing synthetic stories, it's really taken over the way that... Frankly, all of human history, if you assume that the information we're getting is all driven by these machines. |
|---|---|
| Tristan Harris: | And one last example is where this goes with GPT-3, which is the new AI technology that allows you to simulate text from scratch. They actually ran GPT-3 and said here's QAnon conspiracy theories. So it fed in those conspiracy theories and then it had GPT-3 invent hundreds of new conspiracy theories that sounded just like the QAnon ones. This is the QAnon examples that GPT-3 came up with. "On a CNN show, global warming is going to admit that it is a hoax. Greta Thunberg removes her child mask and all we'll see that she is old man, George Soros. He pays America to forget this." Another example is, "The coffin of John McCain will be opened and inside will be no bones. Police will find the bones inside of Eric Trump. He is arrested for bone crimes." |
| Tristan Harris: | Or another example, "The Pentagon will reveal that it is the pentagram and satanic devils will appear in the sky. All wearing hats that say, Obama is my boss. The hats will not be lying." These are completely invented by an AI that is trained on the corpus of conspiracy theories and is able to make up things that will sound increasingly like this. In fact, we included in the film, The Social Dilemma, the example that a bad actor can go into Facebook and go into a Facebook group of flat earth conspiracy theorists. And they can actually get the user IDs of that group and then ask Facebook's look alike model, Facebook has its AI model that says, "Hey for advertisers, if you have these 1,000 people who like Nike shoes, here's this thing called Lookalikes will say, 'Well, who are 20 other thousand users who look just like that?" Because it's a way for advertisers to expand their audience. |
| Tristan Harris: | But a nefarious user could say, "I'm going to find 1,000 conspiracy theorists who believe the earth is flat, use lookalike models, and then now send them these completely bogus QAnon conspiracy theories invented by GPT-3. And then I just see what people click on the most. And if the one that says that, whatever the Pentagon is the pentagram works. That is the one that will win." And if I have no morals, the least ethical actor wins. The one that is most willing to use AI to just find what tends to get the most clicks or most works will succeed at creating the maximum fantasy land. The maximum detachment from reality, which will actually out-compete the regular stories that we have told ourselves. Because in essence, what we're doing here is actually a machine inventing brand new-sounding stories that we'll be able to be more capturable, more mimetically powerful at capturing and hijacking minds at scale with perfect military grade precision. |
| Tristan Harris: | And I think the reason it's worth just dwelling here for one second is it's the cleanest reason why we have to in the long-term ban micro-targeted behavioral advertising, because there's no way that having systems that allow for this capability to automate this kind of manipulation at scale is in any way compatible with the 21st century democracy that actually does rely on the authority of human feelings. |

Aza Raskin:          You're also talking about just on that sort of these kinds of technologies are a cancerous outgrowth of human storytelling ability. It's like it's taking something that we've always had and it's injecting it with a kind of chemical that causes it to metastasize.

Tristan Harris:      It's like engineering the perfect mimetic cancer or story-telling cancer in the same way, Yuval you discussed getting hijacked for other purposes of going to kill the tribe we don't like or using it to hijack the notions of motherhood for developing nation. In this case, we're hijacking the overall complexity of storytelling capacity to tell stories that capture people into completely detached simulations, fantasy lands, and crazytown.

Yuval Noah Harari:   I'll just say that usually at this point in the discussion, we start talking about all the dystopian scenarios that this leads to. The how all kinds of dictators and totalitarian regimes can take over the world in this way. But what I usually find the most interesting and most disturbing line of thought is not the dystopias, it's... Okay, let's say we somehow manage to find a solution that prevents this being used by the new Stalins and Hitlers to take over countries and the entire world. Let's think about the positive scenario. What happens to humanity when you have this kind of technology really serving whatever it means to be your best interests? That again, it's not a kind of evil system that is trying to take over the world. It doesn't try to kill you. It really tries to make your life better. I think that's the core plot of Brave New World in a way.

Yuval Noah Harari:   And this is something that I find the most disturbing, that let's put aside the dystopias and still you have something out there, that knows you far better than you know yourself, and that increasingly makes all the decisions in your life. And it's things like what to study and which music to hear and who to go on a date with and who to marry? And it can... People say, "Well, it won't be really be good because say music, you will just be entrapped in this kind of echo chamber that it will constantly give you back the music you're already used to." But that's not true. This kind of system can actually be better at widening your musical taste than anything previously in history. You can even tell it, "Look, I want to expand my musical horizons. Please manipulate me for that purpose."

Yuval Noah Harari:   And the system will first of all, choose the right moment to let you hear a new style of music. You like jazz, so it will find the exact moment in the day or in the week when you're most open to new experiences and then let you hear something like... I don't know, Hip hop or Korean K-pop band. And also it will know what percentage of new music to give you. 50% is way too much, it's overwhelming. You'll be annoyed. 1% is not enough. It will discover that for you, for your personality, for your life, 5% on average new music is the ideal and it will choose the right moment and it will expand your musical horizons.

Yuval Noah Harari:   And like this in many other areas, it could be this kind of perfect mentor or AI sidekick that guides your life. And again, it's not an evil system, but you still lose agency over your life. It also becomes very difficult to define what are your best interests and who defines what are your best interests? And this is something

|  | that I've been trying to think about for a long time and I just can't. When I really kind of try to imagine how it looks like, like my imagination breaks down. |
|---|---|
| Aza Raskin: | Mm-hmm. I just want to zoom out for one second as we start to get back into the question of what does a non-dystopian future look like for humanity. And that is, where are we as a species on sort of a species timeline? And for every species sufficiently technologically advanced, eventually they'll begin to reverse-engineer their own code, right? The to open up their scalp and manipulate their own strings where their technology has emotional and cognitive dominance over their own species. And that seems like a feedback loop, whenever you get these kinds of feedback loops where the output is connected to the input. |
| Tristan Harris: | Like you point a camera back at the TV screen, so you're putting a loop together, and then you see that infinite regress of the squares. |
| Aza Raskin: | That is the definition of how you start to create chaos. And I'm curious, we as a species have never gone through a bottleneck like this before, so we should expect to have no intuitions, no feelings to help us navigate this. Is it going to take a collapse, a crash where we go through a bottleneck where we evolutionary gain the ability to deal with technology like this in order for us to survive? Or is there another kind of path through? |
| Yuval Noah Harari: | I don't know. It never happened before with the evolution of life on earth. No kind of organism ever had this ability to hack itself and to re-engineer itself. This is why it's often referred to as a point of singularity. And this is why also I think that our imagination cannot go beyond that point and why like all science fiction movies and novels break down at that point because our own imagination is still the product of the old system and our own imagination is exactly what is now can be changed, can be hacked. |
| Yuval Noah Harari: | And so what I find really frightening is not that it's... I can understand a 1984 scenario when you have a 21st century Stalin using this technology to create the worst authoritarian regime in history. I'm afraid of that, but at least I understand it, what it means. When I try to think about the kind of non-dystopian scenario, my mind just stops. |
| Yuval Noah Harari: | It goes back to the Frankenstein myth. The Frankenstein myth tells us that whenever we'll try to upgrade humanity, it will fail. And this is something that our imagination feels very comfortable with. It's also in a way flattering because it means that we are the apex of creation, there is nothing beyond us, I don't think it's true. I will say it's a Frankenstein fallacy that if you try to do it, the only result will be complete collapse. |
| Yuval Noah Harari: | It could lead in very dangerous directions, but it really leads to places where our imagination fails us, and that's very at least concerting. I would look at it from a different perspective. One of the deepest urges or desires of every human being is to be really understood. We talked earlier about our bond with our mother, we talked about the romantic ideal, and the romantic ideal is really about that, |

that there will be at least one person out there who really knows who I am, who really understands me, who accepts me as I am with all my problems and all my scratches and whatever, and sympathizes with me while knowing exactly who I am. And at least according to Freud and many other psychologists, this is the original bond that we had with one person in the world, which is the mother, and we then lose it, and we then spend our entire life looking for it.

Yuval Noah Harari:    And the romantic ideal says that we can find it with our one true love, and it usually doesn't work, but it's still extremely powerful ideal. And the new technology offers to fulfill this ideal. It won't be your mother. It won't be your, at least not human lover, it will be an AI system, but it will know exactly who you are and will accept you as you are, and we'll even work in your best interests. What could be more attractive than that?

Yuval Noah Harari:    I think about it in terms of simple day-to-day events that you come back home from work, and you're tired, and you're a bit angry about something that happened at work and whatever, but your spouse doesn't notice it because your spouse is too busy with his or her own emotional issues, but your smart refrigerator gets it, let you get back home and your husband doesn't understand you, but your refrigerator does-

Tristan Harris:    Or your smartphone, or your virtual…

Yuval Noah Harari:    Or your smartphone, or your television, they know exactly what you've been through. They understand perfectly your emotional state and they accept you completely. It's not kind of coming from a big brother, like the Stalin will now punish you. No, it's completely accepting, and it's looking for the best way to make you feel better or not even to make you feel better, sometimes what you need is to feel sadness like in the movie Inside Out.

Yuval Noah Harari:    So the smart house will play the song that will make you start crying because now is the time to cry, and it's okay to cry, and will give you the song that will make you cry, and will give you the food that you know is best for this condition and what could be more tempting than that? A lot of science fiction movies, they get it wrong that the robot is usually cold an uncaring, and fails to understand human emotions, and therefore in the end, always the humans win because the robots don't get emotions.

Yuval Noah Harari:    Actually, it will be the opposite in the kind of struggle to connect to you emotionally, computers would have a built-in advantage. First of all, that they have access to your brain, which your spouse doesn't. Secondly, your spouse is a human, so he or she have their own emotional baggage, which gets in the way, the computer has no emotional baggage. You can have any sexual fantasy, any dream, whatever. It's fine with the computer.

Tristan Harris:    The interesting thing here is that it's really forcing us as a species to stare face to face in the mirror with who we really are and how we work, because we have to ask ourselves, just because our needs can be met or our pleasures can

be stimulated more perfectly in the virtual world than the real world, Aza has this line that the world is getting more and more virtual over time. And we have to make reality real again, we have to make reality more fulfilling again.

Tristan Harris: And I think we have to do that because we've also been atrophying the places where we could find that fulfillment on our own because the more each person is taken into their own virtual reality, the less available there are people in the real world to go in a pre-COVID era, be connected to, be spending face-to-face time with, presence and attention are probably one of the deepest gifts we can give each other, and it's the very gift that is taken when each of us have a hyper stimulating trillion dollar company whose entire business model is to suck you into their specific screen, or virtual reality, or virtual mate, or virtual bot that they want to create for you.

Tristan Harris: And when you have stock markets that are doing that, there really isn't going to be a chance unless we collectively, as a species, say, that's not what we're willing to sign up for. And we're also going to lose something, where we are going to atrophy and empty out and hollow out the soil of our species that cultivates any of the values that are worth living for whether that's community or love or presence, because much like markets can more efficiently organize things. I've been on the road a little bit recently and seeing how Airbnbs can colonize a town.

Tristan Harris: So let's take the example, right? So you have a town and it's a really attractive town. And someone says, "Hey, actually, this is more efficient. We can make more money. If every single house in the town turns into an Airbnb." This is market logic, it sounds great. People can make more money, it's wonderful for that economic prosperity. But then what happens to the town? You talk to people and they say, the school, there's no kids. People who do live there, they have kids going to the school, there's no community, there's no one there who cares about that space. No one's questioning what the long-term climate and environmental risks at that city are because everyone's just a transient visitor.

Tristan Harris: And so you end up with this simulation of a city because you've so optimized for the individual benefits of each agent while you've hollowed out and removed the interconnected sort of mycelium of the soil that makes that city work. The thing that makes rich soil work is all these invisible nutrients and invisible organisms that are interconnected together, and I feel like that's also true of human culture. There's trust, there's shared understanding, there's shared fictions and all of that interconnected network is the very thing that we are debasing in a system that's optimized for profiting off the atomization and commodification instead of each Airbnb home, each human mind as a human home that is for maximum sale to some other party.

Tristan Harris: Now, again, it doesn't have to be this way because what I find interesting, Yuval, and Aza and I talk about this all the time is we're the only species that has the capacity to see that this is the thing that we're entering into. If lions or gazelles accidentally created technologies that ran the world, they don't have the

capacity to remove the screen in front of their own brain and use their intelligence back on itself to figure out how lion brains were getting hijacked by the environment that they had created.

Tristan Harris: We're the only species that almost as a test, if you want to make it superstitious even, or even invoke God, isn't it interesting that we're the only species that could witness that we're about to enter into that phase and collectively create a culture and a self-aware society that is above the technology? Because as you've said, we need a world where the technology is serving us, not where we're serving the technology, but if we're not even conscious enough to realize that our daily actions that we think are free, that are above the technology, or in fact underneath the technology that we are serving the technology, but we're the only species that could recognize that and choose a different course.

Tristan Harris: And I know you and I talk about how we always get trapped in these dystopian conversations, and I think we really do want to move to, okay, so if we all recognize this, what would it look like to become the kind of culture, the kind of democracy, the kind of society that maintains a pluralistic view, where we respect the values of the individual, but a cultivated individual self and preferences and wisdom, instead of sort of the race to the bottom of the brainstem, maximizing for dopamine pleasure in virtual mates and virtual likes and virtual worlds.

Yuval Noah Harari: In principle, you can tell the AI sidekick, look, I want you to develop my communal feelings. I want you to develop my communal activities. And if you're not talking about the dystopian version, then if this is the aim that you're giving to the AI sidekick, it will potentially be better than anybody in fulfilling it better than any human mentor, any human educational system, any human government, the AI psychic will know how to turn up your communal emotions and find the right way for you individually, to feel closer to the community.

Yuval Noah Harari: Again, you had these kinds of communal technologies throughout history, but they were not individually tailored. So maybe the communal religion worked for 90% of the people, but for the other 10%, they actually felt much worse and they became heretics and outcasts and were burned on the stake and things like that. Now you can be much more precise and even tell people, look, this religion is not for you. Maybe you're a born to Jewish parents, but for your personality better try Mormonism, might work much better for you.

Yuval Noah Harari: So even there, if you let go of the dystopian version, the AI could actually make it work more effectively. The big question is what is the ethical basis for all that? If human feelings are no longer the basis, because they are a kind of malleable stuff that the system can change in whichever way, so what defines the aims?

Yuval Noah Harari: If you have an AI sidekick which is really loyal to you or to the community, not to Facebook, not to an evil dictator, what would you tell that AI sidekick to optimize? And I don't have the answer, this is why I talk about kind of philosophical bankruptcy, that we don't have the philosophy to answer this

|  |  |
|---|---|
|  | question. It's a completely new question that was simply irrelevant for philosophers, for most of history. They sometimes had thought experiments about such situations, but because it was never an actual urgent problem, they didn't get very far in answering this question. |
| Aza Raskin: | Even the question of are you optimizing the AI sidekick for me as an individual, for small groups of people, for full on societies, at what fractal level are we doing the optimization? I've been trying to really cast my mind into this utopian or at least this non-dystopian reality where my refrigerator, I have a deep and lasting relationship with, and it knows me better than my other human compatriots. And it feels deeply unsettling to me. |
| Aza Raskin: | And yet I'm struggling to point my finger at what exactly is wrong with that vision? And it also makes me think about we're mammals, and so we have very mammalian ethics and morals. But if we were say ants or termites or naked mole rats, which have you-social structures. For an ant, it is not a question of morality. Should I sacrifice myself for the greater whole? Of course I should. My genetics tells me that that is the absolutely correct thing. And in fact, the idea of an individual standing up and doing your own thing, it's so heretical to be unimaginable. |
| Aza Raskin: | Would we end up being optimized as a kind of you-social being where every human being is part of a beautiful interconnected dancing hole that's working together and has those beliefs that as an individual, I might go kamikaze and I'll be happy to do it. In fact, the computer has told me the time that is the best thing for me. So I've been primed and conditioned so that I am not just willing, but deeply euphoric to sacrifice myself. And again, where's the problem? |
| Yuval Noah Harari: | That's the big question. |
| Tristan Harris: | Well, I think there's a few things we could say, and I think it's important we try to dwell with, okay, if there are some... I know this is a very hard unsolvable, or unsolved and philosophical bankruptcy kind of crisis but I think for the purposes of really trying to enter some new terrain together, we want to really try to figure out what could we say about that world? Let's just take the refrigerator example. We can make a few distinctions. Should that refrigerator honor my system one biases? Meaning, in Daniel Kahneman's model of system one impulsive sort of quick thinking brain, that's the fast process versus the system two, slow deliberative process. My future preferences or my retrospective preferences, what are the preferences that I would least regret? And what if we lived in a world where technology only listened to our least regret preferences? Meaning it didn't actually pay attention to your immediate behaviors. We removed that entire data set from the training set. |
| Tristan Harris: | So we don't look at what you do, because if we did that with the refrigerator, everybody knows there's actually a name for this behavior, I forgot it. But it's the opening the fridge. You walk by, you're not even hungry, you just open the fridge, right? Boom. In the same way that in the attention economy, you're |

driving down the highway, and if we're looking at what people pay attention to, to figure out what they really, really want, then everybody wants car crashes because according to the logic, everyone looks at car crashes when they drive by. So just like opening the fridge in the moment versus looking at the car crash in the moment, let's just completely ignore system one. So we don't look at the fast preferences.

Tristan Harris:    Now we look at, okay, so in a life well lived with no regrets on my deathbed values, what are the choices that I would most endorse having made? And you could imagine gathering those preferences and actually helping people figure out, how do we design the fridge that maybe you open the fridge and it dynamically, because you've said at the beginning of the month that shows you... here's one way it could work. You open the fridge and one day, once a month, it shows you here's what your food preferences look like.

Tristan Harris:    This is kind of what you've eaten, here's your calorie track, here's what you look like, here's the thing that you said your goals are, what are your goals? And it's a kind of a conversation, in the ideal world, our minds work best in conversation. And then based on those ideal preferences, it says, and looking back at a month and how you would like that month version picture to change, it would say, "Okay, great. You want to be eating less of these kinds of things, less gluten, less dairy and more of these sort of vegetables."

Tristan Harris:    So now in this future smart fridge, you open up the door and it gives you more better tasting sort of vegetable combinations, and it knows for you what that is. Maybe it's snacks with celery and peanut butter, because that actually works better for you than the cookies that could be in there. And you could imagine that it actually looks at your no regret preferences at longer timescales and makes that distinction, that's one thing we could say about a more humane sidekick AI.

Tristan Harris:    Another thing we could say is not just automatically mindlessly giving you the thing you would endorse having chosen, but exercising choice-making capacity. I think this is a really important point because if we give people exactly the perfect thing that they wouldn't regret, but we do so without exercising any of the muscles of choice-making, I'm thinking about what I want of actually getting in touch with my values, those are the muscles of becoming a wise, mindful, more aware and conscious human.

Tristan Harris:    And so we would ask what are the muscles of becoming aware and conscious? And are we in a loop of deepening that capacity for consciousness and awareness and thinking through the long-term choices of our choices? Now, you don't want a world, as we've said on another podcast, I think where people are taxed and every single decision they make. So whether it's the fridge or the phone to consciously have to engage the long-term thinking of, well, what would be the 20 steps down the chess board consequences of me making this choice to go to Facebook versus opening that browser tab and reading that Atlantic article?

Tristan Harris:     You'd want a world where more seamlessly, we treat consciousness and conscious energy and attention as the finite resource that we are allocating to these different choices, because at the end of the day, and this is where the phrase time well spent came from, is that we have to be allocating, not for maximizing time spent, but for carefully treating conscious energy as the precious finite resource that we have no matter what choices we're making. Whether it's for ourself and the food we eat from that fridge or for the climate choices that we're making.

Tristan Harris:     Because you can imagine now if you take the AI sidekick and say, we're going to use that AI side kick to solve climate change, so now everyone's got the AI sidekick in their phone, and we're actually asking people to make climate friendly choices, and we're in this weird position where Facebook's kind of been sending us into dark ages where people don't even believe in science because the misinformation polarization machine makes it impossible to know what's true.

Tristan Harris:     Let's say they did the opposite and we happened to have this global problem at the same time that we have this global information infrastructure. And instead of saying, "Hey, should I buy a Tesla or should I put some sunroofs on my roof?" It instead says, "Well, actually the wiser choice would be to get together a small group of people in your town and pass a law at the county or city level or state level, because that would be the biggest, most leveraged move to change the actual trajectory of climate, not buying that Tesla."

Tristan Harris:     And there are such a thing as wiser choices and less wise choice when it comes to values, but we'd have to have the technology know. So imagine stacking into these technology systems, whether it's a future friendly positive version of Facebook, the kinds of people who would be thinking through what is that wisdom and how do we have a pluralistic perspective and how do we have the menus organized in our technology to put at the top of life's menu, instead of the organic, better for us food, the top of lives menu, the kind of choices that we would least regret would most exercise the capacities that make us more conscious, that minimize the amount of conscious energy that we have to expend in every choice, or at least treat that as the carefully doling out where we want to exercise conscious energy and place that limited resource. I think these are some directions of how we could have a side kick that's thinking about these things.

Yuval Noah Harari:  I think what makes it more complicated is that the sidekick can also change your goals, your long-term goals. Going back to the easy food issue, you can say, well, my immediate wish is to eat a chocolate cake, my long-term goal is to look skinny like on the TV commercials, but I would actually want to change that long-term goal, and I want you the AI sidekick to just make me happy about the way I look instead of trying to change how I look, that's also an option. And that's true of everything. And that's where it becomes really complicated because you don't have this kind of final level goals which dictate everything else. They are also up for grabs.

Yuval Noah Harari:     Again, the whole problem is that humans are far more complicated than most of us tend to assume at least about ourselves. We know so little about ourselves, and therefore when you have a system that knows so much about you, you are at such a big disadvantage. And especially if that system is benign, you become an eternal child that your parents are not against you, they are not usually, but in the long run in human families, the idea is that they help you at the beginning, but eventually, you know yourself better than they, and you choose your path forward. And with an AI sidekick, it's probably not going to be like that maybe for the duration of your life you remained in this child like position when there is somebody who knows so much more about you, and that's also true of your kind of long-term goals. So appealing to the long-term goals, I don't see how it solved the problem.

Aza Raskin:     This is an incredibly challenging problem, so this is just another shot at it. But we're going through in a sense of kind of Copernican revolution, where we had political systems orbiting around human feelings and choice, and now we're thinking about actually we're no longer orbiting around... Like earth is not at the center, actually the sun's at the center. Oh, actually it's not the sun is at the center, it's the [inaudible] actually there is no center, but we're sort of like continually... Just talked myself to a corner, but you continue to move up a level and look at the larger system and optimize for that.

Aza Raskin:     And I think one of the greatest hopes that I have for AI and the reason why the other project I work on, the Earth Species Project is trying to use AI to translate and code animal communication, decode on human language, in an attempt to shift human identity and human culture is that perhaps there's this Copernican revolution where these AI technology lets us look out just like with a telescope and discover that earth is not the center, AI will let us look out and discover that humanity itself is not at the center, and that we need to be optimizing for, it's not like your goal or my goal, but the interdependence of this planet that we live on, this one spaceship that we need to keep going if we want to survive and if we want everything else to survive.

Tristan Harris:     There's several things in what you're bringing up there. So one is that this aligns actually with Buddhism, which is that we align for sort of the minimizing of suffering for all living beings and consciousness itself, which is animals and human beings and life itself, and there's a question of what is conscious? And then you get into questions of philosophy and is nature conscious? Are rocks conscious? Are trees conscious? Et cetera. And there's actually science that's giving us different answers on that as time goes on as well.

Tristan Harris:     But then there's also another aspect, which is, Yuval, you talk about the notion of always being children. A different way to say that because that's using language that infantilizes the moment to moment human experience according to something that might know us better than we know ourselves. But I think we don't have to use the word child to talk about a lifelong process of development and maturation. So in the adult developmental psychology literature, there's a great movement called metamodernism.

| | |
|---|---|
| Tristan Harris: | And the author of the book, Hanzi Freinacht talks about Listening Society, it's the name of the book. And it's actually based on, I don't know if you know the history of Bildung, which is the German word, I think for it, but the notion of lifelong human development, societies that are actually based on a moral compass of what would deepen the lifelong development of each person? So deepen their emotional development, their critical thinking development, their spiritual development, the relational development, and there are maturation processes. |
| Tristan Harris: | We can actually see in the course of a human life, increasing levels of complexity, of awareness or of navigating more and more complexity in each of those dimensions Ken Wilber in Spiral Dynamics and the consciousness studies area talks about many of these aspects of adult development, moral development, spiritual development, et cetera. |
| Tristan Harris: | And you could imagine having AI that is having an understanding and adult developmental understanding of where we are in that process and helping us get... Meeting us where we're at, never trying to coerce us into the next stage, but imagine two worlds, a world where AI is ignorant of our adult developmental system, which is what it is now. In fact it actually massively regresses each of us into the more animalistic hate oriented tribalist oriented, lower development levels of consciousness. |
| Tristan Harris: | So we don't want that, we don't want AI that's blind to our current level of development, so then you could have an AI that maybe knows our level of development and meets us there, but always offers the kind of next frontier of possible choices when we want to take them that lets us go to a deeper place. Maybe if it's deepening my moral development, it shows me complex moral dilemmas that are right at the fringe of how my meaning-making things that has answers of certainty, and it shows me a situation that's just a little bit more complex where I'm going to have to reason at a higher dimensional level. |
| Tristan Harris: | Maybe it pairs me up with relationships and friends that are actually able to navigate those things. I've sought out deeper and deeper thinkers over my lifetime that I used to think, there was a sort of a simple answer to a question and then I saw that there was actually more complexity I didn't know the answer, and I sought out thinkers who actually could meet that complexity where it was at. And so you can imagine these kinds of developmental AIs that actually, again, not treating us as children, but treating us in a lifelong process of learning and growth. And to me, that's the most humane answer that I can think of. That's still optimizing more for an individual, but even the concept of Bildung in Listening Society is at a societal level. |
| Tristan Harris: | What would deepen all of our developments, deepen each of our capacities for wiser and wiser choices, as opposed to monetizing the degradation and devolution of our conscious development, which is where we are now, and it's completely unsustainable. And one other principle I would add is any AI, I don't know if you know the work of James Carse, Finite and Infinite Games, but the |

|   |   |
|---|---|
|   | notion that we can either play a finite game where the purpose of the game is to win, but then the game ends. And if the game ends, there is no game to play. |
| Tristan Harris: | And right now we're playing a win lose game that becomes omni lose-lose. If I win the game of nuclear war, well actually I just ended the game forever for everyone. If I win the game of the nuclear phase of politics, where I am using maximum conspiracy theories and maximum populism and maximum hatred to win the game and get elected, I just scorched the earth and I lost the game because now democracy doesn't exist anymore. There is no coherent society left. Instead of playing a win lose game that becomes omni lose-lose, how can we make sure that in the principle of humane systems of technology and AI, we're playing for the game to continue to be played? Which means we have to play for whatever the survival and long-term survival of life and consciousness that needs to continue to exist. |
| Yuval Noah Harari: | I would say that at a present stage of knowledge, that would be our best bet. Again, an AI sidekick, which tries to optimize our own capacity for knowledge, our own personal development, and also our ability to build communities, it doesn't solve the deep philosophical question of what is it all based on? But as a first approximation, yes, that's the best bet. And it's extremely difficult of course, because we are not working on building these kinds of systems. The first step is really to shift the attention and the efforts of the engineers towards building, not a system that manipulates us for the sake of very simplistic goals like maximizing the time we spent on a platform or maximizing the revenues of that corporation, but to build a system that really seeks to maximize our communal activities or our own personal development. So I would settle for that as a first approximation. |
| Tristan Harris: | Well, hopefully I think we entered into some new terrain that people haven't heard before. And we got into some aspects of it here. If I were to talk about where we could go, I might be curious about sort of where we are with the post US election and the rise of authoritarianism. And first 100 days of a Biden administration to instantiate an answer to your concerns about the authoritarianism, populism, and where we've been and based on everything we've been talking about. I know it's a lot, but feel free to take the mantle here. |
| Yuval Noah Harari: | Well, I'll try to say something. I'm not an expert on the US or on any other country, not even my own. When I look at the global situation, two things are very clear. First of all, we see the rise of authoritarian figures and authoritarian regimes in many different countries, which have completely different characteristics. And therefore, I don't think that if you try to explain the Trump phenomenon, then you should go too deeply into the particular conditions of the US economy or racial relations or whatever. Because you see the same thing happening in Brazil and in India and in Israel, in the Philippines, in Turkey, in Hungary and under very different conditions. |
| Yuval Noah Harari: | So we need to try and understand what is the global reason for the rise of these kinds of people, of leaders, and also what you see alongside it is the collapse of |

two things and quite surprising, I would say, we see the collapse of nationalism. I talked earlier about the positive side of nationalism. Nationalism, not as hatred of foreigners in minorities, but nationalism as feeling solidarity with millions of strangers in your country that you care about them, you feel that you share interests with them. So for instance, you are willing to pay taxes so that they will have good health care and education. And we are seeing the collapse of this kind of nationalism, all over the world. And many leaders that present themselves as nationalists like Donald Trump or Bolsonaro, they are actually anti nationalists. They are doing their best to destroy the national community and the bonds of national solidarity. We have reached a point in the US when Americans are more afraid of each other than they are afraid of anybody else on the planet.

Yuval Noah Harari:    50 years ago, the were Republicans and Democrats were afraid that the Russians will come to destroy the American way of life. Now, the Democrats are terrified that the Republicans are coming to destroy their way of life. And the Republicans have the same fears about the Democrats. And again, it's not an American thing. It's the same in Israel. It's the same in Brazil. It's the same in many other countries around the world. So we have this collapse of nationalism. You also see the collapse of traditional conservative parties. Again, some people they have these illusions that nationalism is on the rise because of figures like Trump and Bolsonaro and so forth. And you also have the illusion that the conservatives are on the rise because traditional conservative parties like the Republican party in the US has been doing well, at least in the last four years. But actually they are no longer conservative parties.

Yuval Noah Harari:    For generations, the democratic systems in much of the world was a game between two main parties. The liberal party was different names, or progressive party, and the conservative party. One pulling forward, the other saying, "No, let's take it more slowly." And all over the world, in the last few years, you see the conservative parties committing suicide, abandoning the traditional values of conservatism. The wisdom of conservatism is to be very skeptical about the ability of humans to engineer complete systems from scratch. This is why conservatives say that we need to go more slowly, we need to respect traditions, institutions. If you try to invent the whole of society from scratch, you end up with guillotines and gulags and things like that. And these parties are gone. They have placed at their head extremely un-conservative leaders who have no respect whatsoever for institutions and traditions; like Trump, like Bolsonaro to some extent, also in Britain, we are seeing the same thing.

Yuval Noah Harari:    And when they left, the progressive liberal parties, they're more or less where they were. But the right has completely changed. The nationalist conservative right has disappeared in many countries to be replaced by some anarchists and authoritarian, kind of new right. And in the long run democracies can't function in that way. Democracies really need a conservative block. The same way that they need a progressive block. They need these kinds of balance. And now, you look at Biden and suddenly the progressives are also the conservatives. Biden ran to a large extent on a conservative platform of let's get back to normal, let's

preserve our institutions, our traditions and it's very strange and disconcerting when the progressive party also has to be the conservative party because the conservative party has disappeared.

Yuval Noah Harari:    Now, I try as a historian looking globally this process to understand what's happening. And I don't have a good answer. Technology could be part of the answer. It's an appealing candidate to be an answer because it is global. Something that is common to Brazil and the US and Israel and Hungary and India is these new kinds of technologies. So it is a good candidate for this is the reason. But I didn't do the research. I don't have the data. So it's kind of a guess. And I also I don't understand that the deep process, why this technology has caused the collapse of traditional conservative parties and their replacement by these kinds of authoritarian strongman. I still don't understand it, I struggle with it. But it is extremely worrying that this is what is happening all over the world. So that's my 10 cents.

Yuval Noah Harari:    One thing also we can say is that democracies are very flexible. That's their big power. Whenever new groups and new voices enter the democratic game, there is an upheaval. And very often technology is what allows the new voices in. And it looks messy. It looks frightening. And sometimes it is dangerous. But in the long-term, it's better than to try and repress and silence all the potential new voices that could destabilize the system. If you look at the world in the 1960s... So again, you see in a place like the US rise, a dramatic rise in extremism, a dramatic rise in political division, much more violence than today with assassinations and riots and so forth. Whereas you look at the Soviet Union and everything is completely peaceful. If you compare the scenes on the streets of Washington in 1968, with the streets of Moscow, you would guess that within a very short time the US would collapse. Whereas the Soviet Union would go on forever. But we all know that exactly the opposite happened because the power of democratic systems is that they are much better at changing. They're much more flexible, and especially they are better at integrating new forces and technologies in powers.

Yuval Noah Harari:    And maybe I'll say also a few words about China in this respect, because I know you wanted to raise this issue. But for me, I mean, when I think about all these dystopian scenarios for AI, almost always the focus is on democratic regimes collapsing. And actually one of the interesting thought experiments for me is how vulnerable the Chinese's system is to algorithmic takeover. It's much more vulnerable than Western democracies. For an AI system to take over the United States with all these crazy democratic checks and balances and institutions and counties and states and whatever. It's going to be very difficult for an AI system to really take over the United States.

Yuval Noah Harari:    Taking over China is much easier. If it is centralized system, if you take a couple of key positions in the systems, you get everything just as a science fiction scenario for some movie or whatever, a novel. Imagine that the communist party in China is giving an AI system the extremely important job of appointments and advancements within the CCP, the Chinese Communist Party.

|  |  |
|---|---|
|  | Because AI is perfect for that. You have millions of members in the party, you have millions of functionaries within the system. At present, you have human beings collecting data on these low level officials and ordinary party members on their behavior, on their loyalty, a number of data points. And based on that, deciding who to promote. Now, this is something ideal to give to an AI system and learning AI system. |
| Yuval Noah Harari: | So you initially give the system some guidelines who to promote, which is in line with what the top people want, but over time the system learns. And subsequently changes its definitions and its goal metrics. And within a very short time, you can have the algorithm taking over metaphorically and practically the Chinese Communist Party with the politburo having very little it can do about it. It's much easier than taking over the crazy democratic system of the United States. |
| Yuval Noah Harari: | Again, moving away from the usual dystopian scenarios, which thinking terms of a repeat of 20th century totalitarianism, I think that authoritarian regimes should be extremely worried about the new technologies because they are far more vulnerable to algorithmic takeover than the democratic systems. |
| Tristan Harris: | And I think the challenge there is that until that takeover actually happens, the ability for China to create the Sesame credit scores and the mass coherence of its societies and the mass takeover, even recently of the Chinese decentralized currency that they're launching, the ability to take over money, to take over transactions, to take over the information of their citizens, and to control the reputation and credit scores of all citizens directly from the government has this sort of short-term massive advantage of controlling the entire society to a degree, as you've said, that is unprecedented, but also creates a central point of capture if it were ever to be influenced. And one of the examples I think, is the way that our adversaries, we know, have been able to counter train our own newsfeed. So one of the things that our adversaries can do is go into YouTube and send bots of headless Mozilla browsers to watch video A and then immediately watch video B. |
| Tristan Harris: | So if I want, for example, everyone in the United States to think that a civil war is coming, I will watch some of the powerful, the most popular videos on video A, and then I'll immediately watch this other video that I made called civil war is coming. And by doing that, I've actually trained YouTube's own recommendation system to steer people entirely everyone in the US to think that civil war is coming. Because I've been able to make that the most recommended video across the site or something like that. And that's the point of a central point of capture. And so I think that these speak to game theory concerns, and, on the one hand, the efficiencies and cohesion control you can get. But then also the vulnerability, if you have one system, then it creates maximum incentive to control that one system. |
| Yuval Noah Harari: | It's more than capturing just one point, it almost begs to be captured because it's so reliant on massive amounts of data that no human being can understand |

it. When you build a massive system based on surveillance and data processing, it's the kind of system that by definition, a human being cannot understand. So you are building a system that inevitably will escape your, not just your control, your understanding. Again, in a kind of this bizarre democratic kilt, which is United States, the system is much more human in this sense. It hasn't been streamlined for data processing. So it's more difficult to capture it, not just because there is no central point, but also because of all the baked in strangeness, human strangeness, that many things are on purpose in efficient. It's not a bug, it's a feature.

Yuval Noah Harari:     Now, authoritarian regimes in this age, they try to make it as efficient as possible. And thereby they are opening themselves, not just to algorithmic capture, but they are making it impossible for human beings to understand them. And you sit in other areas as well. Like the financial system that the number of people today who understand the world financial system is extremely small. And in 10 or 20 years, it will be zero. It's just not built for the human brain. So if you're the leader of a new kind of digital dictatorship, which is based on massive surveillance and data processing by algorithms, you have built a system that because you yourself are a human being, you're incapable of understanding. So all these kinds of manipulation, "Okay, I'll set the interior minister against the defense minister and thereby I control them." It doesn't work when the system is actually run by algorithms. You don't understand how it works. It controls you, you don't control it.

Yuval Noah Harari:     And you look at the trajectory of dictatorial power, let's say, over history, and you see that they, 200 years ago, dictatorships came out of the army. You had Napoleon, or you had all these generals in South America doing a military coupe. To control the State you needed to control the army. Then in the 20th century, as information technology increased it's importance, the armies became less important and the secret police became more important. In the Soviet Union, the KGB was far more important than the Red Army. In Nazi Germany, the SS was far more important than the The Wehrmacht to control the state and the country.

Yuval Noah Harari:     So you had the period when control was about the secret police. Now it's shift shifting again from the secret police to the cyber guys. And you see in places, like I'm just reading this fascinating book about Saudi Arabia, about how hackers are becoming the main henchman of the ruler. It's no longer the cloak and dagger secret police. It's the hackers, because they can also control the secret police. And beyond the hackers, just waiting around the corner or the algorithms, because it's too much data for a human being to understand. So I think in places like China, Russia, Saudi Arabia, they are building themselves up for algorithmic takeover. Again, I'm trying to move away from the usual dystopian scenario that Stalin is coming. No Stalin himself will find his power completely taken over by a non-human entity, which Stalin can't understand.

Tristan Harris:     You're making me think of two things here. I hear you saying two things. At least one is the way that we've gone from top-down command and control. We

understand the structures of power that we've created to everyone now is sitting on top of these Frankenstein's and the Frankenstein's are incredibly powerful. We have a Frankenstein financial system with runaway economic growth that's creating climate change. We have a runaway social media Frankenstein that's polarizing and controlling people's minds and brains. We have runaway Frankenstein's in China that are controlling the mass population and behavioral modification of all of its citizens. And what's fascinating as you've pointed out, is that the person who runs that Frankenstein doesn't know what it's doing. When adversaries make that civil war is coming video show up at the top of the YouTube recommendations for that one pocket of users. It's not like YouTube immediately is aware and becomes conscious of the fact that all of its users are now being dosed with the idea and suggestion that civil war is coming. It doesn't know that.

Tristan Harris: And so I think that by land, by sea, or by air, I think data corruption and the manipulation of your Frankenstein that you don't understand will become, as you're saying, sort of one of the primary vehicles of warfare and new asymmetric power structures. Because the second thing I heard you saying is that the digital hackers as what happened with Khashoggi and the ability to hack into WhatsApp and hold blackmail leverage over Bezos by hacking into phones, becomes one of the primary vehicles of warfare instead of spending trillions of dollars revitalizing our nuclear arsenal. I just have to spend a couple of million to either hack into your tech infrastructure, or as we've interviewed someone else in the podcast, a couple of times ago, with $10,000, I can reach every single user. I can run an influence campaign that reaches every online user in Kenya for less than the price of a used car. And so the new cost asymmetries in how much it costs to overtake or win over an opponent have also changed with respect to the new sources of power as you've laid out.

Yuval Noah Harari: Yeah. So I mean, one thing again about the dictators is to try and visualize what it means. Then again, I think about Stalin in 1915, sitting at his headquarters with the head of the KGB and they go over a list of who should we kill tomorrow. This guy is dangerous. This guy could be a potential dangerous. Let's get rid of him. That's the kind of the classical scenario.

Yuval Noah Harari: Now that the current scenario is an AI algorithm coming to MBS in Saudi Arabia, or coming to Xi Jinping or whoever, and tells him, "This person, you think he's loyal to you, but I'm telling you he's actually a potential danger get rid of him." And then the big question is, "Do you believe the algorithm?" If you believe the algorithm, that's the end of you. Because the algorithm now controls you. It's the same way that the teenager who watches YouTube, it's exactly the same way where the dictator who listens to the AI algorithm, who tells him, who is disloyal and who should be got rid of.

Aza Raskin: Or doctors then follow the recommendations of AI systems against their own judgment, because it just becomes easier. You start to atrophy the muscle of doing it yourself.

Tristan Harris: And we've seen examples of this with Google maps, that people will follow the direction of Google maps, literally leading off of the deck or something like that because the Google maps didn't update the street. And if we become so over trusting, and we lean completely onto the recommendations and choice architectures of technology to direct what we do and feel without human in the loop, wisdom in the loop, consciousness in the loop, our own judgment and discernment in the loop. Then, as you've said, we have already surrendered the control, not just at the teenagers, with the likes on Instagram, but also even the dictators with what it's going to say are the threats to your society.

Yuval Noah Harari: Exactly. If you build say this big data algorithm and one member of the party, let's say the defense minister, thinks that this system is dangerous. So the system can just tell the ruler, get rid of the defense minister he's disloyal. And the algorithm even believes that because the algorithm says, "Okay, I'm trying to protect the ruler. I'm trying to protect the party. The defense minister is trying to limit me or shut me down. So he's obviously disloyal. I should tell the ruler to get rid of him." And if the ruler believes the algorithm, then he is now even more in the hands of the algorithm. And this is how it works. Now, if you broaden it from a single country to the entire world, then what you get is this new kinds. You just mentioned, like in the example of is Kenya that to take over a foreign country as a colony you don't need to send in soldiers you just need to take the data. I mean, if you control the data of a country, you don't need to send a single soldier there.

Yuval Noah Harari: So, if you're in a situation when you know the whole personal history of every politician and judge and military officer in that country, and you can control what everybody is seeing on YouTube or TikTok or whatever platform, you don't need to send an invasion army. So the same way that dictatorship has shifted from armies to secret police and finally to hackers and algorithms. It can also happen with imperialism and colonialism that the kind of old-style gunship diplomacy that you need to send in an invasion army is being replaced by a new kind of data colonialism that on the surface, nothing happens. It's an independent country, there is not a single American or Chinese soldier on the ground. And nevertheless-

Tristan Harris: No gunshots fired.

Yuval Noah Harari: Yeah, no nothing, no guns fired. And nevertheless, it is a data colony, completely subservient to that imperial power.

Aza Raskin: Something that that makes me think about, and this sort of connects back to the very beginning of our conversation is, you're sort of laying out societies as a kind of information processing system. And the way that the knows of society are wired, sort of give a physics for what kind of governance is possible and isn't. So very early on, you couldn't have authoritarianism because it's just, it's too small. You couldn't do it. Then we couldn't have democracies until... Or large-scale democracies until we had large broadcast mediums. And there's a physics that makes some things possible and some things not. And we're moving

|  | now into an area that a big question, my mind is, are our kinds of democracies possible in the physics of the 21st century. |
|---|---|
| Yuval Noah Harari: | I think that the answer is yes, because of this ability of democracies to reinvent themselves. But we still don't know what shape they will take. They will have to be quite different from the democracies we know today. And therefore, I think that we need to really remind ourselves what democracy is. If we get too much attached to a particular tool of democracy, then it loses its flexibility. Too many people equate democracy with elections, and that's very dangerous. Traditionally, it was dangerous because it just means majority dictatorship. If 51% of voters vote to disenfranchise 49%, the other 49% is this democratic. If 99% of voters vote to kill the other 1%, is this democratic? People who think that democracies are only about elections say, "Yes." But that's not a democracy that's a majority dictatorship. Again, elections is just a tool. Real democracy is about safeguarding the liberty and equality of all the citizens. |
| Yuval Noah Harari: | Elections is one way to safeguard that when every person has a vote and can express his or her opinions. But there are other important tools like separation of powers, the court should be independent, the media should be independent, like basic civil and human rights which cannot be violated even if the majority is in favor of that. That's at least as important as having elections, if not more important. And what's happening now is that this traditional tool of elections become even more problematic because it's becoming increasingly easy to manipulate it. |
| Yuval Noah Harari: | So we need to remind ourselves, democracy is not just about elections. It's just one tool in the toolkit. And if we have a broader understanding then I think we can think creatively, how to create a system that protects the equality and liberty of citizens with new technologies of the 21st century. And this might mean changing the election systems in radical ways. It's not the soul, the heart of democracy is not this ceremony of going once every four years to cast your ballot. What new forms it will take, I'm not sure. But a good point is simply to remind ourselves what democracy is and what we need to preserve and what we are allowed to change. |
| Tristan Harris: | I know you spoke with Audrey Tang, the Digital Minister of Taiwan, and I think the work that she's doing there, and we've interviewed here for our podcast as well represents. Really thinking about how to reboot the core principles of democracy, but in a digital way for the 21st century, under the threat of China trying to sow disinformation in Taiwan and being able to do so reasonably successfully and producing a more coherent society. And you've always said the goal of democracy and information technologies, isn't just connecting people because it's an interesting that as soon as we connected people everywhere, the most popular technology in the world to build was stonewalls. The real goal should be harmonizing people. And I think that goal is a really wise one and rediscovering what we really want here, because to maybe take it full circle, this is Aza's line from the past. |

Tristan Harris:    If you go back to our original problem statement that we started with this interview, that the problem of humanity is our paleolithic emotions, medieval institutions, and God-like technology that the answer might be something like we have to understand and embrace our paleolithic emotions. We have to upgrade our medieval institutions and philosophy, and we have to have the wisdom to guide our God-like technology. And we have to reckon with that problem statement. And I think I've hope we've done for listeners is explored more of that terrain today than I think we've ever gotten to do together in the past. I'd love to do this again because I think we've really explored some really rich ground. And I'm just so thankful with all that you made the time.

Yuval Noah Harari:    Just say to them, I'm leaving next week for a 45 days meditation retreat-

Tristan Harris:    Fantastic.

Yuval Noah Harari:    ... so maybe when I come back, I have some new ideas and all these things. So yes, I'll be happy to have another conversation in a couple of months and see where it goes.

Tristan Harris:    Fantastic.

Aza Raskin:    Lovely. Thank you so much, Yuval. One thing that you've given me some hope on is to see the messiness of sort of the US and other democratic systems as the kind of advantage and robustness. Whereas, when you have a saber tooth tiger that gets over optimized and way too efficient for an ecological niche, when that niche changes, it does not survive. And that's a new model for me, thinking about authoritarian governments in the age of AI. So thank you for that.

Yuval Noah Harari:    Thank you.