NICHOLAS THOMPSON
BUSINESS   10.04.2018 07:00 AM

When Tech Knows You Better Than You Know Yourself

Historian Yuval Noah Harari and ethicist Tristan Harris discuss the future of artificial intelligence with WIRED editor in chief Nicholas Thompson.

WHEN YOU ARE 2 years old, your mother knows more about you than you know yourself. As you get older, you begin to understand things about your mind that even she doesn't know. But then, says Yuval Noah Harari, another competitor joins the race: "You have this corporation or government running after you, and they are way past your mother, and they are at your back." Amazon will soon know when you need lightbulbs right before they burn out. YouTube knows how to keep you staring at the screen long past when it's in your interest to stop. An advertiser in the future might know your sexual preferences before they are clear to you. (And they'll certainly know them before you've told your mother.)

I had the chance to speak with Harari, the author of three best-selling books, and Tristan Harris, who runs the Center for Humane Technology and who has played a substantial role in making "time well spent" perhaps the most-debated phrase in Silicon Valley in 2018. They are two of the smartest people in the world of tech, and each

spoke eloquently about self-knowledge and how humans can make themselves harder to hack. As Harari said, "We are now facing not just a technological crisis but a philosophical crisis."

Please read or watch the entire thing. This transcript has been edited for clarity.

Nicholas Thompson: Tristan, tell me a little bit about what you do and then Yuval, you tell me too.

Tristan Harris: I am a director of the Center for Humane Technology where we focus on realigning technology with a clear-eyed model of human nature. And I was before that a design ethicist at Google where I studied the ethics of human persuasion.

Pre-Cyber Week Starts Now. Pre-Cyber Week Starts Now.
Get WIRED for $10 $5.

Subscribe Now

Yuval Noah Harari: I'm a historian and I try to understand where humanity is coming from and where we are heading.

NT: Let's start by hearing about how you guys met because I know that goes back a while. When did the two of you first meet?

YNH: Funnily enough, on an expedition to Antarctica, we were invited by the Chilean government to the Congress of the Future to talk about the future of humankind and one part of the Congress was an expedition to the Chilean base in Antarctica to see global warming with our own eyes. It was still very cold and there were so many interesting people on this expedition

TH: A lot of philosophers and Nobel laureates. And I think we particularly connected with Michael Sandel, who is a really amazing philosopher of moral philosophy.

NT: It's almost like a reality show. I would have loved to see the whole thing. You write about different things, you talk about different things but there are a lot of similarities. And one of the key themes is the notion that our minds don't work the way that we sometimes think they do. We don't have as much agency over our minds as perhaps we believed until now. Tristan, why don't you start talking about that and then Yuval jump in, and we'll go from there.

TH: Yeah I actually learned a lot of this from one of Yuval's early talks where he talks about democracy as "Where should we put authority in a society?" And where we should put it in the opinions and feelings of people.

Yuval Noah Harari, left, and Tristan Harris with WIRED editor in chief Nicholas Thompson. WIRED

But my whole background: I actually spent the last 10 years studying persuasion, starting when I was a magician as a kid, where you learn that there are things that work on all human minds. It doesn't matter whether they have a PhD, whether they're a nuclear physicist, what age they are. It's not like, Oh, if you speak Japanese I can't do this trick on you, it's not going to work. It works on everybody. So somehow there's this discipline which is about universal exploits on all human minds. And then I was at the Persuasive Technology Lab at Stanford that teaches engineering students how you apply the principles of persuasion to technology. Could technology be hacking human feelings, attitudes, beliefs, behaviors to keep people engaged with products? And I think that's the thing that we both share is that the human mind is not the total secure enclave root of authority that we think it is, and if we want to treat it that way we're going to have to understand what needs to be protected first.

YNH: I think that we are now facing really, not just a technological crisis, but a philosophical crisis. Because we have built our society, certainly liberal democracy with elections and the free market and so forth, on philosophical ideas from the 18th century which are simply incompatible not just with the scientific findings of the 21st century but above all with the technology we now have at our disposal. Our society is built on the ideas that the voter knows best, that the customer is always right, that ultimate authority is, as Tristan said, is with the feelings of human beings and this assumes that human feelings and human choices are these sacred arena which cannot be hacked, which cannot be manipulated. Ultimately, my choices, my desires reflect my free will and nobody can access that or touch that. And this was never true. But we didn't pay a very high cost for believing in this myth in the 19th and 20th century because nobody had a technology to actually do it. Now, people—some

people—corporations, governments are gaming the technology to hack human beings. Maybe the most important fact about living in the 21st century is that we are now hackable animals.

Hacking a Human

NT: Explain what it means to hack a human being and why what can be done now is different from what could be done 100 years ago.

YNH: To hack a human being is to understand what's happening inside you on the level of the body, of the brain, of the mind, so that you can predict what people will do. You can understand how they feel and you can, of course, once you understand and predict, you can usually also manipulate and control and even replace. And of course it can't be done perfectly and it was possible to do it to some extent also a century ago. But the difference in the level is significant. I would say that the real key is whether somebody can understand you better than you understand yourself. The algorithms that are trying to hack us, they will never be perfect. There is no such thing as understanding perfectly everything or predicting everything. You don't need perfect, you just need to be better than the average human being.

NT: And are we there now? Or are you worried that we're about to get there?

YNH: I think Tristan might be able to answer where we are right now better than me, but I guess that if we are not there now, we are approaching very very fast.

TH: I think a good example of this is YouTube. You open up that YouTube video your friend sends you after your lunch break. You come back to your computer and you think OK, I know those other times I end up watching two or three videos and I end up getting sucked into it, but this time it's going to be really different. I'm just going to watch this one video and then somehow, that's not what happens. You wake up from a trance three hours later and you say, "What the hell just happened?" And it's because you didn't realize you had a supercomputer pointed at your brain. So when you open up that video you're activating Google's billions of dollars of computing power and they've looked at what has ever gotten 2 billion human animals to click on another video. And it knows way more about what's going to be the perfect chess move to play against your mind. If you think of your mind as a chessboard, and you think you know the perfect move to play—I'll just watch this one video. But you can only see so many moves ahead on the chessboard. But the computer sees your mind and it says, "No, no, no. I've played a billion simulations of this chess game before on these other human animals watching YouTube," and it's going to win. Think about when Garry Kasparov loses against Deep Blue. Garry Kasparov can see so many moves ahead on the chessboard. But he can't see beyond a certain point like a mouse can see so many moves ahead in a maze, but a human can see way more moves ahead and then Garry can see even more moves ahead. But when Garry loses against IBM Deep Blue, that's checkmate against humanity for all time because he was the best human chess player. So it's not that we're completely losing human agency and you walk in to YouTube and it always addicts you for the rest of your life and you never leave the screen. But everywhere you turn on the internet there's

basically a supercomputer pointing at your brain, playing chess against your mind, and it's going to win a lot more often than not.

"Everywhere you turn on the internet there's basically a supercomputer pointing at your brain, playing chess against your mind, and it's going to win a lot more often than not."

— TRISTAN HARRIS

NT: Let's talk about that metaphor because chess is a game with a winner and a loser. But YouTube is also going to—I hope, please, Gods of YouTube—recommend this particular video to people, which I hope will be elucidating and illuminating. So is chess really the right metaphor? A game with a winner and a loser.

TH: Well the question is, What really is the game that's being played? So if the game being played was, Hey Nick, go meditate in a room for two hours and then come back to me and tell me what you really want right now in your life. And if YouTube is using 2 billion human animals to calculate based on everybody who's ever wanted to learn how to play ukulele it can say, "Here's the perfect video I have to teach later that can be great." The problem is it doesn't actually care about what you want, it just cares about what will keep you next on the screen. The thing that works best at keeping a teenage girl watching a dieting video on YouTube the longest is to say here's an anorexia video. If you airdrop a person on a video about the news of 9/11, just a fact-based news video, the video that plays next is the Alex Jones InfoWars video.

NT: So what happens to this conversation?

TH: Yeah, I guess it's really going to depend! The other problem is that you can also kind of hack these things, and so there are governments who actually can manipulate the way the recommendation system works. And so Yuval said like these systems are kind of out of control and algorithms are kind of running where 2 billion people spend their time. Seventy percent of what people watch on YouTube is driven by recommendations from the algorithm. People think that what you're watching on YouTube is a choice. People are sitting there, they sit there, they think, and then they choose. But that's not true. Seventy percent of what people are watching is the recommended videos on the right hand side, which means 70 percent of 1.9 billion users, that's more than the number of followers of Islam, about the number followers of Christianity, of what they're looking at on YouTube for 60 minutes a day—that's the average time people spend on YouTube. So you got 60 minutes, and 70 percent is populated by a computer. The machine is out of control. Because if you thought 9/11 conspiracy theories were bad in English, try 9/11 conspiracies in Burmese and Sri Lanka and in Arabic. It's kind of a digital Frankenstein that's pulling on all these levers and steering people in all these different directions.

NT: And, Yuval, we got into this point by you saying that this scares you for democracy. And it makes you worry whether democracy can survive, or I believe you say the phrase you use in your book is: Democracy will become a puppet show. Explain that.

YNH: Yeah, I mean, if it doesn't adapt to these new realities, it will become just an emotional puppet show. If you

go on with this illusion that human choice cannot be hacked, cannot be manipulated, and we can just trust it completely, and this is the source of all authority, then very soon you end up with an emotional puppet show.

And this is one of the greatest dangers that we are facing and it really is the result of a kind of philosophical impoverishment of just taking for granted philosophical ideas from the 18th century and not updating them with the findings of science. And it's very difficult because you go to people—people don't want to hear this message that they are hackable animals, that their choices, their desires, their understanding of who am I, what are my most authentic aspirations, these can actually be hacked and manipulated. To put it briefly, my amygdala may be working for Putin. I don't want to know this. I don't want to believe that. No, I'm a free agent. If I'm afraid of something, this is because of me. Not because somebody planted this fear in my mind. If I choose something, this is my free will, And who are you to tell me anything else?

NT: Well I'm hoping that Putin will soon be working for my amygdala, but that's a side project I have going. But it seems inevitable, from what you wrote in your first book, that we would reach this point, where human minds would be hackable and where computers and machines and AI would have better understandings of us. But it's certainly not inevitable that it would lead us to negative outcomes—to 9/11 conspiracy theories and a broken democracy. So have we reached the point of no return? How do we avoid the point of no return if we haven't reached there? And what are the key decision points along the way?

YNH: Well nothing is inevitable in that. I mean the technology itself is going to develop. You can't just stop all research in AI and you can't stop all research in biotech. And the two go together. I think that AI gets too much attention now, and we should put equal emphasis on what's happening on the biotech front because in order to hack human beings, you need biology and some of the most important tools and insights, they are not coming from computer science, they are coming from brain science. And many of the people who design all these amazing algorithms, they have a background in psychology and brain science because this is what you're trying to hack. But what should we realize? We can use the technology in many different ways. I mean for example we now are using AI mainly in order to surveil individuals in the service of corporations and governments. But it can be flipped to the opposite direction. We can use the same surveillance systems to control the government in the service of individuals, to monitor, for example, government officials that they are not corrupt. The technology is willing to do that. The question is whether we're willing to develop the necessary tools to do it.

"To put it briefly, my amygdala may be working for Putin."

– YUVAL NOAH HARARI

TH: I Think one of Yuval's major points here is that the biotech lets you understand by hooking up a sensor to someone features about that person that they won't know about themselves, and they're increasingly reverse-engineering the human animal. One of the interesting things that I've been following is also the ways you can ascertain those signals without an invasive sensor. And we were talking about this a second ago. There's something called Eulerian Video magnification where you point a computer camera at a person's face. Then if I put a supercomputer behind the camera, I can actually run a

mathematical equation, and I can find the micro pulses of blood to your face that I as a human can't see but that the computer can see, so I can pick up your heart rate. What does that let me do? I can pick up your stress level because heart rate variability gives me your stress level. I can point—there's a woman named Poppy Crum who gave a TED talk this year about the end of the <u>poker face</u>, that we had this idea that there can be a poker face, we can actually hide our emotions from other people. But this talk is about the erosion of that, that we can point a camera at your eyes and see when your eyes dilate, which actually detects cognitive strains—when you're having a hard time understanding something or an easy time understanding something. We can continually adjust this based on your heart rate, your eye dilation. You know, one of the things with <u>Cambridge Analytica</u> is the idea—you know, which is all about the hacking of Brexit and Russia and all the other US elections—that was based on, if I know your big five personality traits, if I know Nick Thompson's personality through his openness, conscientiousness, extrovertedness, agreeableness, and neuroticism, that gives me your personality. And based on your personality, I can tune a political message to be perfect for you. Now the whole scandal there was that Facebook let go of this data to be stolen by a researcher who used to have people fill in questionnaires to figure out what are Nick's big five personality traits? But now there's a woman named Gloria Mark at UC Irvine who has done research showing you can actually get people's big five personality traits just by their click patterns alone, with 80 percent accuracy. So again, the end of the poker face, the end of the hidden parts of your personality. We're going to be able to point AIs at human animals and figure out more and more signals from them including their micro expressions, when you smirk and all these things, we've got face ID cameras on all of these phones. So now if you have a tight loop where I can adjust the political messages in real time to your heart rate and to your eye dilation and to your political personality. That's not a world that you want to live in. It's a kind of dystopia.

YNH: In many contexts, you can use that. It can be used in class to figure if one of the students is not getting the message, if the student is bored, which could be a very good thing. It could be used by lawyers, like you negotiate a deal and if I can read what's behind your poker face, and you can't, that's a tremendous advantage for me. So it can be done in a diplomatic setting, like two prime ministers are meeting to resolve the Israeli-Palestinian conflict, and one of them has an ear bug and a computer is whispering in his ear what is the true emotional state. What's happening in the brain in the mind of the person on the other side of the table. And what happens when the two sides have this? And you have kind of an arms race. And we just have absolutely no idea how to handle these things. I gave a personal example when I talked about this in Davos. So I talked about my entire approach to this to these issues is shaped by my experience of coming out. That I realized that I was gay when I was 21, and ever since then I was haunted by this thought. What was I doing for the previous five or six years? I mean, how is it possible? I'm not talking about something small that you don't know about yourself—everybody has something you don't know about yourself. But how can you possibly not know *this* about yourself? And then the next thought is a computer and an algorithm could have told me that when I was 14 so easily just by something as simple as following the focus of my eyes. Like, I don't know, I walk on the beach or even watch television, and there is—what was in the 1980s, *Baywatch* or something—and there is a guy in a swimsuit and there is a girl in a swimsuit and which way my eyes are going. It's as simple as that. And then I think, What would my life have been like, first, if I

knew when I was 14? Secondly, if I got this information from an algorithm? I mean, if there is something incredibly, like, deflating for the ego, that this is the source of this wisdom about myself, an algorithm that followed my movements?

Coke Versus Pepsi

NT: And there's an even creepier element, which you write about in your book: What if Coca-Cola had figured it out first and was selling you Coke with shirtless men, when you didn't even know you were gay?

YNH: Right, exactly! Coca-Cola versus Pepsi: Coca-Cola knows this about me and shows me a commercial with a shirtless man; Pepsi doesn't know this about me because they are not using these sophisticated algorithms. They go with the normal commercials with the girl in the bikini. And naturally enough, I buy Coca-Cola, and I don't even know why. Next morning when I go to the supermarket I buy Coca-Cola, and I think this is my free choice. I chose Coke. But no, I was hacked.

NT: And so this is inevitable.

TH: This is the whole issue. This is everything that we're talking about. And how do you trust something that can pull these signals off of you? If a system is asymmetric—if you know more about me than I know about myself, we usually have a name for that in law. So, for example, when you deal with a lawyer, you hand over your very personal details to a lawyer so they can help you. But then they have this knowledge of the law and they know about your vulnerable information, so they could exploit you with that. Imagine a lawyer who took all of that personal information and sold it to somebody else. But they're governed by a different relationship, which is the fiduciary relationship. They can lose their license if they don't actually serve your interest. And similarly a doctor or a psychotherapist. They also have it. So there's this big question of how do we hand over information about us, and say, "I want you to use that to help me." So on whose authority can I guarantee that you're going to help me?

YNH: With the lawyer, there is this formal setting. OK, I hire you to be my lawyer, this is my information. And we know this. But I'm just walking down the street, there is a camera looking at me. I don't even know it's happening.

TH: That's the most duplicitous part. If you want to know what Facebook is, imagine a priest in a confession booth, and they listened to 2 billion people's confessions. But they also watch you around your whole day, what you click on, which ads of Coca-Cola or Pepsi, or the shirtless man and the shirtless women, and all your conversations that you have with everybody else in your life—because they have Facebook Messenger, they have that data too—but imagine that this priest in a confession booth, their entire business model, is to sell access to the confession booth to another party. So someone else can manipulate you. Because that's the only way that this priest makes money in this case. So they don't make money any other way.

NT: There's large corporations that will have this data, you mentioned Facebook, and there will be governments. Which do you worry about more?

"To put it briefly, my amygdala may be working for Putin," says Yuval Noah Harari. WIRED

YNH: It's the same. I mean, once you reach beyond a certain point, it doesn't matter how you call it. This is the entity that actually rules, whoever has this kind of data. I mean, even if in a setting where you still have a formal government, and this data is in the hands of some corporation, then the corporation if it wants can decide who wins the next elections. So it's not really that much of a choice. I mean there is choice. We can design a different political and economic system in order to prevent this immense concentration of data and power in the hands of either governments or corporations that use it without being accountable and without being transparent about what they are doing. I mean the message is not OK. It's over. Humankind is in the dustbin of history.

NT: That's not the message.

YNH: No that's not the message.

NT: Phew. Eyes have stopped dilating, let's keep this going.

YNH: The real question is we need to get people to understand this is real. This is happening. There are things we can do. And like, you know, you have the midterm elections in a couple of months. So in every debate, every time a candidate goes to meet the potential voters, in person or on television, ask them this question: What is your plan? What is your take on this issue? What are you going to do if we are going to elect you? If they say "I don't know

what you're talking about," that's a big problem.

TH: I think the problem is most of them have no idea what we're talking about. And that's one of the issues is I think policymakers, as we've seen, are not very educated on these issues.

NT: They're doing better. They're doing so much better this year than last year. Watching the Senate hearings, the last hearings with Jack Dorsey and Sheryl Sandberg, versus watching the Zuckerberg hearings or watching the Colin Stretch hearings, there's been improvement.

TH: It's true. There's much more, though. I think these issues just open up a whole space of possibility. We don't even know yet the kinds of things we're going to be able to predict. Like we've mentioned a few examples that we know about. But if you have a secret way of knowing something about a person by pointing a camera at them and AI, why would you publish that? So there's lots of things that can be known about us that to manipulate us right now that we don't even know about. And how do we start to regulate that? I think that the relationship we want to govern is when a supercomputer is pointed at you, that relationship needs to be protected and governed by a set of laws.

User, Protect Thyself

NT: And so there are three elements in that relationship. There is the supercomputer: What does it do? What does it not do? There's the dynamic of how it's pointed. What are the rules over what it can collect, what are the rules for what it can't collect and what it can store? And there's you. How do you train yourself to act? How do you train yourself to have self-awareness? So let's talk about all three of those areas maybe starting with the person. What should the person do in the future to survive better in this dynamic?

TH: One thing I would say about that is, I think self-awareness is important. It's important that people know the thing we're talking about and they realize that we can be hacked. But it's not a solution. You have millions of years of evolution that guide your mind to make certain judgments and conclusions. A good example of this is if I put on a VR helmet, and now suddenly I'm in a space where there's a ledge, I'm at the edge of a cliff, I consciously know I'm sitting here in a room with Yuval and Nick, I know that consciously. So I've let the self-awareness—I know I'm being manipulated. But if you push me, I'm going to not want to fall, right? Because I have millions of years of evolution that tell me you're pushing me off of a ledge. So in the same way you can say—Dan Ariely makes this joke actually, a behavioral economist—that flattery works on us even if I tell you I'm making it up. It's like, Nick I love your jacket right now. I feel it's a great jacket on you. It's a really amazing jacket.

NT: I actually picked it out because I knew from studying your carbon dioxide exhalation yesterday...

TH: Exactly, we're manipulating each other now...

The point is that even if you know that I'm just making that up, it still actually feels good. The flattery feels good.

And so it's important we have to sort of think of this as like a new era, a kind of a new enlightenment where we have to see ourselves in a very different way and that doesn't mean that that's the whole answer. It's just the first step we have to all walk around—

NT: So the first step is recognizing that we're all vulnerable, hackable.

TH: Right, vulnerable.

NT: But there are differences. Yuval is less hackable than I am because he meditates two hours a day and doesn't use a smartphone. I'm super hackable. So what are the other things that a human can do to be less hackable?

YNH: So you need to get to know yourself as best as you can. It's not a perfect solution but somebody is running after you, you run as fast as you can. I mean it's a competition. So who knows you best in the world? So when you are 2 years old it's your mother. Eventually you hope to reach a stage in life when you know yourself even better than your mother. And then suddenly, you have this corporation or government running after you, and they are way past your mother, and they are at your back. They're about to get to you—this is the critical moment. They know you better than you know yourself. So run away, run a little faster. And there are many ways you can run faster, meaning getting to know yourself a bit better. So meditation is one way. And there are hundreds of techniques of meditation, different ways work with different people. You can go to therapy, you can use art, you can use sports, whatever. Whatever works for you. But it's now becoming much more important than ever before. You know it's the oldest advice in the book: Know yourself. But in the past you did not have competition. If you lived in ancient Athens and Socrates came along and said, "Know yourself. It's a good, it's good for you." And you said, "No I'm too busy. I have this olive grove I have to deal with—I don't have time." So OK, you didn't get to know yourself better, but there was nobody else who was competing with you. Now you have serious competition. So you need to get to know yourself better. But this is like the first maxim. Secondly as an individual, if we talk about what's happening to society you should realize you can't do much by yourself. Join an organization. Like if you're really concerned about this, this week join some organization. Fifty people who work together are a far more powerful force than 50 individuals, who each of them is an activist. It's good to be an activist. It's much better to be a member of an organization. And then there are other tested and tried methods of politics. We need to go back to this messy thing of making political regulations and choices of all this. It's maybe the most important. Politics is about power. And this is where power is right now.

"The system in itself can do amazing things for us. We just need to turn it around, that it serves our interests, whatever that is and not the interests of the corporation or the government."

— YUVAL NOAH HARARI

TH: I'll add to that. I think there's a temptation to say OK, how can we protect ourselves? And when this conversation shifts into my smartphone not hacking me, you get things like, Oh I'll set my phone to grayscale. Oh I'll turn off notifications. But what that misses is that you live inside of a social fabric. We walk outside. My life depends on the quality of other people's thoughts, beliefs, and lives. So if everyone around me believes a conspiracy theory

because YouTube is taking 1.9 billion human animals and tilting the playing field so everyone watches Infowars—by the way, YouTube has driven 15 billion recommendations of Alex Jones InfoWars, and that's recommendation. And then 2 billion views—if only one in a thousand people believe those 2 billion views, that's still 2 million people.

YNH: Mathematics is not our strong suit.

TH: And so if that's 2 million people, that's still 2 million new conspiracy theorists. If you say hey I am a kid I'm a teenager and I don't want to care about the number of likes I get so I'm going to stop using Snapchat or Instagram. I don't want to be hacked from my self-worth in terms of likes. I can say I don't want to use those things but I still live in a social fabric where all my other sexual opportunities, social opportunities, homework transmission, where people talk about that stuff, if they only use Instagram, I have to participate in that social fabric. So I think we have to elevate the conversation from, How do I make sure I'm not hacked. It's not just an individual conversation. We want society to not be hacked, which goes to the political point and sort of how do we politically mobilize as a group to change the whole industry. I mean, for me, I think about the tech industry.

NT: So that's sort of step one in this three step question. What can individuals do? Know yourself, make society more resilient, make society less able to be hacked. What about the transmission between the supercomputer and the human? What are the rules and how should we think about how to limit the ability of the supercomputer to hack you?

YNH: That's a big one.

TH: That's a big question.

NT: That's why we're here!

YNH: In essence, I think that we need to come to terms with the fact that we can't prevent it completely. And it's not because of the AI, it's because of the biology. It's just the type of of animals that we are, and the type of knowledge that now we have about the human body, about the human brain. We have reached a point when this is really inevitable. And you don't even need a biometric sensor, you can just use a camera in order to tell what is my blood pressure. What's happening now, and through that, what's happening to me emotionally. So I would say we need to reconceptualize completely our world. And this is why I began by saying that we suffer from philosophical impoverishment, that we are still running on the ideas of the basically the 18th century, which are good for two or three centuries, which were very good, but which are simply not adequate to understanding what's happening right now. And which is why I also think that, you know, with all the talk about the job market and what should they study today that will be relevant to the job market in 20-30 years, I think philosophy is one of the best bets maybe.

NT: I sometimes joke my wife studied philosophy and dance in college, which at the time seemed like the two

worst professions because you can't really get a job in either. But now they're are like the last two things that will get replaced by robots.

TH: I think that Yuval is right. I think, often, this conversation usually makes people conclude that there's nothing about human choice or the human mind's feelings that's worth respecting. And I don't think that is the point. I think the point is that we need a new kind of philosophy that acknowledges a certain kind of thinking or cognitive process or conceptual process or social process that we we want that. Like for example [James] Fishkin is a professor at Stanford who's done work on deliberative democracy and shown that if you get a random sample of people in a hotel room for two days and you have experts come in and brief them about a bunch of things, they change their minds about issues, they go from being polarized to less polarized, they can come to more agreement. And there's a sort of a process there that you can put in a bin and say, that's a social cognitive sense-making process that we might want to be sampling from that one as opposed to an alienated lonely individual who's been shown photos of their friends having fun without them all day, and then we're hitting them with Russian ads. We probably don't want to be sampling a signal from that person—not that we don't want from that person, but we don't want that process to be the basis of how we make collective decisions. So I think, you know, we're still stuck in a mind-body meatsuit, we're not getting out of it, so we better learn how do we use it in a way that brings out the higher angels of our nature, the more reflective parts of ourselves. So I think what technology designers need to do is ask that question. So a good example, just to make it practical, let's take YouTube again. So what's the difference between a teenager—let's take the example of: You watch the ukulele video. It's a very common thing on YouTube. There's lots of ukulele videos, how to play ukulele. What's going on in that moment when it recommends other ukulele videos? Well there's actually a value of someone wants to learn how to play the ukulele, but the computer doesn't know that, it's just recommending more ukulele videos. But if it really knew that about you instead of just saying, here's like infinite more ukulele videos to watch, it might say and here's your 10 friends who know how to play ukulele that you didn't know know how to play ukulele and you can go hang out with them. It could basically put those choices at the top of life's menu.

YNH: The system in itself can do amazing things for us. We just need to turn it around, that it serves our interests, whatever that is and not the interests of the corporation or the government. OK, now that we realize that our brains can be hacked, we need an antivirus for the brain, just as we have one for the computer. And it can work on the basis of the same technology. Let's say you have an AI sidekick who monitors you all the time, 24 hours a day. What do you write? What do you see? Everything. But this AI is serving you as this fiduciary responsibility. And it gets to know your weaknesses, and by knowing your weaknesses it can protect you against other agents trying to hack you and to exploit your weaknesses. So if you have a weakness for funny cat videos and you spend an enormous amount of time, an inordinate amount of time, just watching—you know it's not very good for you, but you just can't stop yourself clicking, then the AI will intervene, and whenever these funny cat videos try to pop up the AI says no no no no. And it will just show you maybe a message that somebody just tried to hack you. Just as you get these messages about somebody just tried to infect your computer with a virus and it can end. I mean, the hardest thing for us is to admit our own weaknesses and biases, and it can go all ways. If you have a bias against

Trump or against Trump supporters, so you would very easily believe any story, however far fetched and ridiculous. So, I don't know, Trump thinks the world is flat. Trump is in favor of killing all the Muslims. You would click on that. This is your bias. And the AI will know that, and it's completely neutral, it doesn't serve any entity out there. It just gets to know your weaknesses and biases and tries to protect you against them.

NT: But how does it learn that it is a weakness and a bias? And not something you genuinely like?

"Everywhere you turn on the internet there's basically a supercomputer pointing at your brain, playing chess against your mind," says Tristan Harris, right. WIRED

TH: This is where I think we need a richer philosophical framework. Because if you have that, then you can make that that understanding. So if a teenager is sitting there and in that moment is watching the dieting video and then they're shown the anorexia video. Imagine instead of a 22-year-old male engineer who went to Stanford, a computer scientist, thinking about, What can I show them that's like the perfect thing? You had a 80-year-old child developmental psychologist who studied under the best child developmental psychologists and thought about in those kinds of moments the thing that's usually going on for a teenager at age 13 is a feeling of insecurity, identity development, like experimentation. What would be best for them? So we think about this is like the whole framework of humane technology is we think this is the thing: We have to hold up the mirror to ourselves to understand our vulnerabilities first. And you design starting from a view of what we're vulnerable to. I think from a

practical perspective I totally agree with this idea of an AI sidekick but if we're imagining like we live in the reality, the scary reality that we're talking about right now. It's not like this is some sci-fi future. This is the actual state. So we're actually thinking about how do we navigate to an actual state of affairs that we want, we probably don't want an AI sidekick to be this kind of optional thing that some people who are rich can afford and other people who don't can't, we probably want it to be baked into the way technology works in the first place, so that it does have a fiduciary responsibility to our best, subtle, compassionate, vulnerable interests.

NT: So we will have government sponsored AI sidekicks? We will have corporations that sell us AI sidekicks but subsidize them, so it's not just the affluent that have really good AI sidekicks?

TH: This is where business model conversation comes in.

YNH: One thing is to change the way that—if you go to university or college and learn computer science then an integral part of the course is to learn about ethics, about the ethics of coding. And it's really, I think it's extremely irresponsible, that you can't finish, you can have a degree in computer science and in coding and you can design all these algorithms that now shape people's lives, and you just don't have any background in thinking ethically and philosophically about what you are doing. You were just thinking in terms of pure technicality or in economic terms. And so this is one thing which kind of makes it into the cake from the first place.

NT: Now let me ask you something that has come up a couple of times that I've been wondering about. So when you were giving the ukulele example you talked about well maybe you should go see friends who play ukulele, you should go visit them offline. And in your book you say that one of the crucial moments for Facebook will come when an engineer realizes that the thing that is better for the person and for community is for them to leave their computer. And then what will Facebook do with that? So it does seem, from a moral perspective, that a platform, if it realizes it would be better for you to go offline and see somebody, they should encourage you to do that. But then, they will lose their money and they will be outcompeted. So how do you actually get to the point where the algorithm, the platform pushes somebody in that direction?

TH: So this is where this business model conversation comes in is so important and also why Apple and Google's role is so important because they are before the business model of all these apps that want to steal your time and maximize attention.

So Android and iOS, not to make this too technical or an industry-focused conversation, but they should theoretically—that layer, you have just the device, who should that be serving? Whose best interest are they serving? Do they want to make the apps as successful as possible? And make the time, you know, the addictive, maximizing, you know, loneliness, alienation, and social comparison all that stuff? Or should that layer be a fiduciary as the AI sidekick to our deepest interests, to our physical embodied lives, to our physical embodied communities. Like we can't escape this instrument, and it turns out that being inside of community and having face to face contact is, you know—there's a reason why solitary confinement is the worst punishment we give

human beings. And we have technology that's basically maximizing isolation because it needs to maximize the time we stay on the screen. So I think one question is how can Apple and Google move their entire businesses to be about embodied local fiduciary responsibility to society. And what we think of as humane technology, that that's the direction that they can go. Facebook could also change its business model to be more about payments and people transacting based on exchanging things which is something they're looking into with the blockchain stuff that they're theoretically working on and also messenger payments. If they move from an advertising-based business model to micropayments, they could actually shift the design of some of those things. And there could be whole teams of engineers at News Feed that are just thinking about what's best for society and then people would still ask these questions of, well who's Facebook to say what's good for society? But you can't get out of that situation because they do shape what 2 billion human animals will think and feel every day.

NT: So this gets me to one of the things I most want to hear your thoughts on, which is Apple and Google have both done this to some degree in the last year. And Facebook has. I believe every executive at every tech company has said "time well spent" at some point in the last year. We've had a huge conversation about it and people have bought 26 trillion of these books. Do you actually think that we are heading in the right direction at this moment because change is happening and people are thinking? Or do you feel like we're still going in the wrong direction?

YNH: I think that in the tech world we are going in the right direction in the sense that people are realizing the stakes. People are realizing the immense power that they have in their hands—I'm talking about people in the tech world—they are realizing the influence they have on politics, on society, on and on. And most of them react, I think, in not the best way possible, but certainly the responsible way in understanding, yes, we have this huge impact on the world. We didn't plan that maybe. But this is happening and we need to think very carefully what to do with that. They don't know what to do with that. Nobody really knows, but at least the first step has been accomplished, of realizing what is happening and taking some responsibility. The place where we see a very negative development is on the global level because all the talk so far has really been kind of internal Silicon Valley, California, USA talk. But things are happening in other countries. I mean all the talk we've had so far relied on what's happening in liberal democracies and in free markets. In some countries, maybe you have got no choice whatsoever. You just have to share all your information and you just have to do what the government sponsored algorithm tells you to do. So it's a completely different conversation and then another kind of complication is the AI arms race that five years ago, or even two years ago, there was no such thing. And now it's maybe the number one priority in many places around the world that there is an arms race going on in AI, and we, our country, needs to win this arms race. And when you enter an arms race situation, then it becomes very quickly a race to the bottom, because you can very often hear this, OK, It's a bad idea to do this, to develop that, but they are doing it and it gives them some advantage and we can't stay behind. We are the good guys. We don't want to do it, but we can't allow the bad guys to be ahead of us so we must do it first. And you ask the other people, they will say the same thing. And this is this is an extremely dangerous development.

TH: It's a prisoner's dilemma. It's a multipolar trap. I mean every actor—no one wants to build slaughter bot

drones. But if I think you might be doing it, even though I don't want to, I have to build and you build it. And we both hold them.

NT: And even at a deeper level if you want to build some ethics into your slaughter bot drones but it will slow you down.

TH: Right. And one of the challenges—one of the things I think we talked about when we first met was the ethics of speed, of clock rate. Because the faster—we're in essence competing on who can go faster to make this stuff but faster means more likely to be dangerous, less likely to be safe. So it's basically we're racing as fast as possible to create the things we should probably be going as slow as possible to create. And I think that you know much like high-frequency trading in the financial markets, you don't want people blowing up whole mountains so they can lay these copper cables so they can trade a microsecond faster. So you're not even competing based on you know an Adam Smith version of what we value or something like that. You're competing based on basically who can blow up mountains and make transactions faster. When you add high-frequency trading to who can program human beings faster and who's more effective at manipulating culture wars across the world, that just becomes this like race to the bottom of the brainstem of total chaos. So I think we have to say how do we slow this down and create a sensible pace, and I think this is also about a humane technology of instead of a child development psychologist, ask someone, like you know the psychologist, what are the clock rates of human decision making where we actually tend to make good thoughtful choices? You probably don't want a whole society revved up to, you know, making 100 choices per hour about something that really matters. So what is the right clock rate? I think we have to actually have technology steer us towards those kinds of decision making processes.

Is the Problem Getting Better, or Worse?

NT: So back to the original question. You're somewhat optimistic about some of the small things that are happening in this very small place? But deeply pessimistic about the complete obliteration of humanity?

TH: I think that Yuval's point is right that, you know there's a question about US tech companies, which are bigger than many governments—Facebook controls 2.2 billion people's thoughts. Mark Zuckerberg is editor in chief of 2.2 billion people's thoughts. But then there's also, you know, world governments or, sorry, national governments that are governed by a different set of rules. I think the tech companies are very, very slowly waking up to this. And so far, you know, with the Time Well Spent stuff, for example, it's let's help people, because they're vulnerable to how much time they spend, set a limit on how much time they spend. But that doesn't tackle any of these bigger issues about how you can program thoughts of a democracy, how mental health and alienation can be rampant among teenagers leading to doubling the rates of teen suicide for girls in the last eight years. So you know we're going to have to have a much more comprehensive view and restructuring of the tech industry to think about what's good for people. And there's going to be an uncomfortable transition, I've used this metaphor, it's like with climate change. There are certain moments in history when an economy is propped up by something we don't want. So the

biggest example of this is slavery in the 1800s. There was a point at which slavery was propping up the entire world economy. You couldn't just say we don't want to do this anymore let's just suck it out of the economy. The whole economy will collapse if you did that. But the British Empire when they decided to abolish slavery, they had to give up 2 percent of their GDP every year for 60 years and they were able to make that transition over a transition period, and so I'm not equating advertising or programming human beings to slavery. I'm not. But there's a similar structure of the entire economy now, if you look at the stock market, like a huge chunk of the value, is driven by these advertising programming–human–animals–based systems. If we wanted to suck out that model, the advertising model, we actually can't afford that transition. But there could be awkward years where you're basically in that long transition path. I think in this moment we have to do it much faster than we've done in other situations because the threats are more urgent.

NT: Yuval, do you agree that that is one of the things we have to think about as we think about trying to fix the world system over the next decades?

YNH: It's one of the things. But again, the problem of the world, of humanity, is not just the advertising model. I mean the basic tools were designed—you had the brightest people in the world, 10 or 20 years ago, cracking this problem of, How do I get people to click on ads? Some of the smartest people ever, this was their job, to solve this problem. And they solved it. And then the methods that they initially used to sell us underwear and sunglasses and vacations in the Caribbean and things like that, they were hijacked and weaponized, and are now used to sell us all kinds of things, including political opinions and entire ideologies. And it's now no longer under the control of the tech giant in Silicon Valley that pioneered these methods. These methods are out there. So even if you get Google and Facebook to completely give it up, the cat is out of the bag. People already know how to do it. And there is an arms race in this arena. So yes, we need to figure out this advertisement business. It's very important. But it won't solve the human problem. And I think now the only really effective way to do it is on the global level. And for that we need global cooperation on regulating AI, regulating the development of AI and of biotechnology, and we are, of course, heading in the opposite direction of global cooperation.

TH: I agree actually that there's this notion of game theory. Sure Facebook and Google could do it, but that doesn't matter because the cat's out of the bag, and governments are going to do it, and other tech companies are going to do it, and Russia's tech infrastructure is going to do it. So how do you stop it from happening?

Not to bring it back—not to equate slavery in a similar way, but when the British Empire decided to abolish slavery and subtract their dependence on that for their economy, they actually were concerned that if we do this, France's economy is still going to be powered by slavery and they're going to soar way past way past us. So from a competition perspective we can't do this but the way they got there was by turning it into a universal global human rights issue. That took a longer time but I think this is like Yuval says, I agree that it's this is a global conversation about human nature and human freedom. If there is such a thing but at least kinds of human freedom that we want to preserve and that I think is something that is actually in everyone's interest and it's not necessarily equal capacity to achieve that because governments are very powerful. But we're going to move in that direction by

having a global conversation about it.

NT: So let's end this with giving some advice to someone who is watching this video. They've just watched an Alex Jones video and the YouTube algorithm has changed and sent him here, and they've somehow got to this point. They're 18 years old, they want to devote their life to making sure that the dynamic between machines and humans does not become exploitive and becomes one in which we continue to live our rich fulfilled lives. What should they do or what advice would you give them?

YNH: I would say that, get to know yourself much better and have as few illusions about yourself as possible. If a desire pops in your mind don't just say well this is my free will. I chose this therefore it's good, I should do it. Explore much deeper. Secondly as I said join an organization. There is very little you can do just as as an individual by yourself. That's the two most important pieces of advice I could give an individual who is watching us now.

TH: And I think your earlier suggestion to understand that the philosophy of simple rational human choice is—we have to move from an 18th century model of how human beings work to a 21st century model of how human beings work. Our work is we're trying to coordinate kind of a global movement towards fixing some of these issues around human technology. And I think that like Yuval says you can't do it alone. It's not like, Let me turn my phone grayscale or let me like you know petition my congressmember by myself. This is a global movement. The good news is no one wants the dystopian endpoint of the stuff that we're talking about. It's not like someone says no, no. I'm really excited about this dystopia, I just want to keep doing what we're doing. No one wants that. So it's really a matter of can we all unify and the thing that we do want and it's it's somewhere in this vicinity of what we're talking about. And no one has to catch the flag but we have to move away from the direction we're going. I think everyone should be on the same page on that.

NT: Well I, you know we started this conversation in a time where the optimistic and I am certainly optimistic that we have covered some of the hardest questions facing humanity that you have offered brilliant insights into them. So thank you for talking and thank you for being here. Thank you. Tristan. Thank you, Yuval.

More Great WIRED Stories

- How to build a floating bridge in 12 minutes
- The new YubiKey will help kill the password
- PHOTOS: Hilariously bored tourists from around the world
- The disastrous science behind disaster preparedness
- We gotta get a better battery. But how?
- Get even more of our inside scoops with our weekly Backchannel newsletter