

# OpenTheo

## Can Robots Become Human? | Rosalind Picard

January 10, 2022



### The Veritas Forum

This program was recorded at a Veritas Forum event on Brown in 2017. The original title was "What does it mean to be human?" and featured Rosalind Picard and Michael Littman. If you enjoyed this episode, please rate, review, and subscribe. And, if you're interested in more content from Veritas, check out our Beyond the Forum podcast. Visit [veritas.org](https://veritas.org) to learn more about the mission of the Veritas Forum and find more resources to explore the ideas that shape our lives.

### Transcript

Hi, this is Carly Regal, the assistant producer of Beyond the Forum, a podcast from the Veritas Forum and PRX. The Forum you're about to listen to is featured in Beyond the Forum second season exploring the intersection between science and God. We interviewed Dr. Roslyn Picard, one of the presenters you're about to listen to, for episode three of our second season.

And we talked with her about her career and artificial intelligence and about how technology can help humans and not just replace them. You can listen to our interview with Ros for Beyond the Forum wherever you listen to podcasts. And you can learn more about the ideas that shape our lives by visiting our website at [veritas.org](https://veritas.org). Thanks for listening and enjoy the Forum.

This is the Veritas Forum podcast, a place for generous dialogue about the ideas that shape our lives. You know, it's interesting how obsessed many people who say they're not interested in religion are in immortality. I know people who claim to be atheists and not interested in religion, but they're actually super interested and they're especially interested in immortality.

This is your host Carly Regal. Today I'm sharing with you a conversation at a Veritas Forum event at Brown University in November 2017. The speakers you will hear from are Dr. Roslyn Picard of MIT and Dr. Michael Lettman of Brown University, as they discuss how artificial intelligence and big data play a part in our day-to-day lives.

You can learn more about the Veritas Forum and events like these by visiting [veritos.org](http://veritos.org). I hope you enjoy their conversation. So we've often heard about the possibility of AI achieving human intelligence. So we ask, is that actually possible? And when will it happen? I think one expert or at least a venture capitalist in the area, Jim Breyer has predicted that not so much predicted, but he's looked at everybody's predictions about when does AI reach human intelligence levels? And he says that the median prediction is 2050.

So not all that long from now. Then there's also a concern that many of us heard about what happens after AI becomes human intelligence levels. Are we going to achieve what's known as a singularity? And what's that going to mean if this does, actually does occur? So we'll wrap today this evening's discussion.

We'll talk about religion, also about religion specifically the Christian faith, and how these sorts of concerns tie into the possibility of AI becoming human, and perhaps more controversially possessing a soul. So we've got two speakers that are going to be talking to us today. First to my right is Ros Picard.

She is the founder and director of the effective computing research group at the MIT Media Lab, co-director of the Media Labs Advancing Wellbeing Initiative, and faculty chair of MIT's Mind, Hand, and Heart Initiative. She has co-founded Empatica, creating wearable sensors and analytics to prove health and affectiva, delivering technology to help measure and communicate emotion. A researcher and inventor with multiple patents, she is the author of over a 250 peer-reviewed scientific articles and has consulted for many companies, including Apple, Samsung, and iRobot.

To my left is my brown colleague, Michael Litman. He's a computer science professor here studying machine learning decision-making under uncertainty. He's the co-director of Brown's Humanity-Centered Robots Initiative, and a fellow of association for the advancement of artificial intelligence, and has earned multiple awards for his teaching and research.

And if he's the astric performer, Michael S. had roles in numerous community theater productions and a TV commercial. So we're going to, before we launch into some of the media questions here, I'm going to ask the panelists to, let's say, a bit more about themselves. So Ros, we'll start with you.

So you are a Christian. How would you describe your current beliefs? Starting with the hard questions here. So I am kind of the secularist and the Christian here tonight.

I'm the token Christian, I suppose. I was not always a Christian. I was an atheist for the first part of my life, and was challenged by people who I actually really admired, who are Christians, that maybe I should learn a little bit more and get a little bit more data to go with my views.

My views at the time were that Christians and actually all religions, I was pretty antagonistic toward, were people who really didn't know their science or maybe they needed a crutch or something I really didn't think they were that smart. Then I started to realize that many of such people were super smart, and they challenged me to read the best-selling book of all time, which is probably still the Bible, the Hebrew and Old Testament and Christian New Testament. And as I was reading that to my, against my desires, I started to change my mind about some things.

And then I thought, "Oh gosh, okay, if this book is influencing me to change my mind toward Christianity or toward belief in God, maybe I should study other world religions." So I started to do that. And as I started learning more and more about different world religions, I meeting people from those religions and going to temples and mosques and others. I started to realize that not only did I have a lot to learn, but I was on a journey that was starting to make me not only believe in God even more, but as I got dragged off to some Christian churches, which I resisted in the beginning, and found somewhere I could ask questions, very important.

I started to realize that the religion was not at all what I thought it was, and that there were some really interesting and very attractive elements that were very historically verified also, not at all what I expected. And as I learned about that, I changed my viewpoint gradually from an atheist to an agnostic to a theist, to somebody who actually believed that the historical Jesus in the New Testament, what's written about him was true. It sounds a little wacky to those who may not come from that background.

It was not an easy process. But as I did that, and then I was challenged to not only believe this, but to put it to practice, that's where things started to really make a difference in my life. And actually, the real reason I'm here right now, spending time talking about something like this as opposed to just my research is because it has made a huge difference in my life, and part of the Christian faith is that there's a gift for everybody in the world, whether you're raised Christian or Hindu or Muslim or Buddhist or atheist or any of a long list of backgrounds.

There's a gift for everybody there, and when I accepted that gift, it made a huge difference in my life for the better, the improvement. So I didn't realize it needed so much improving at the time. It was around me, saw the difference.

And today it is my source of strength, an amazing source of peace and joy and wisdom. And as I think when we build machines and build computers with affective abilities and robots, I often think of the analogy of one who is very wise, giving us instructions and giving us guidance and being there when we don't know what to do. So I find that still is powerful in my work today.

Thank you. Michael. So what are your current beliefs and did any religious beliefs affect Europe? Yeah, so I was born into a Jewish family, but a non-practicing Jewish family, and

so I guess I was a non-practicing atheist because we didn't talk about it at all in any way, one way or the other, so I was non-practicing anything.

I suppose that continued for a long time, then we had children, and there was sort of a, okay, well we have a liberal Jewish mother and a sort of non-practicing Jewish father, how are we going to raise the kids? And so we had to figure out what we believed as a family, and we found something that was super helpful for us is there's a movement called humanistic Judaism, which is sort of like you can be Jewish and you can do holidays and stuff, but there is no particular belief system tied up with that beyond humanism, besides the idea that we as people have responsibility to try to make the world better for other people. And so that ended up being a really good fit. And so I guess that's been our belief system for some time.

I did have this moment though when I was a faculty member at Rutgers, one of my more cantankerous colleagues said to me, "So what's your belief?" I'm like, "Well, I guess I'm agnostic." And he's like, "Basically, you can't be agnostic. You have to either choose. You have to be an atheist or you have to believe in something.

Agnostic is just a wimpy way to say it. I'm like, "Oh, okay, all right, well then I guess I'm atheist." So that was kind of my conversion to atheism. (Laughter) I can't say that it's been a huge source of peace and wisdom, but it is at least... (Laughter) At least I feel like I'm maybe being more honest with myself.

All right, thank you. Okay, let's get to some of the meat of our discussion. So Michael, we'll start with you.

I think most of us at least at this point in time can tell the difference between a human being and a robot. Okay, but this may be tougher in the future. So can we come some sort of defining terms about what makes us a human being, what makes something else a robot? Sure, yeah, yeah, I think that's actually... It's actually really interesting.

I think it brings up a couple things that probably we should lay out early on. Some research in robotics is about... So computer scientists, roboticists in some sense are dualists, right? There's people who work on the software, like the brain, and then there's people who work on the bodies, the physical robots. And the technology for producing bodies that at least from a distance seem very human has been improving substantially, and now you can make very human-looking devices.

But they're just the devices unless they have some kind of software that's driving their behavior that we recognize as being, okay, this is an animate agent. This is something that's human-ish. So there's research on that, and then there's separately research on the software side of trying to decide how do you make a software system that can make decisions and can make judgments and decide what to do.

And they do proceed somewhat independently of each other, but occasionally people try to put them together in various ways and have a robot that has a very human appearance, and at the same time has some very human-like behavior in them. And so from that perspective, the lines start to blur. But I think it's important to keep in mind that just because something looks human, like I don't think that we would necessarily perceive it as human, right? We're very sensitive.

We, all of us, are very sensitive when we're looking at something that's not human and it's trying to be human. If it's not quite right, we often will have a very visceral reaction to that. And some people have speculated that this is actually a way of keeping us away from dead bodies, like it's an evolutionary thing to keep us away from dead bodies.

Like if something looks human but it's really off a little bit, it may be something that you want to give a wide berth to. And so this is one of the explanations that people give for what's called the uncanny valley. Thanks very much.

Yeah, so this sort of idea that we often have almost a disgust reaction to things that are not quite human in their behavior. And part of what you can see when you think about that is that we're actually incredibly well-attuned to what are the signals that an animate agent send us that we recognize as being human. We're very discerning.

And you can get it almost right and not quite right and that's disturbing. I think we're still not at a point where we can make systems that really push all the right buttons to fool us into thinking that's real. You can fool some of the people some of the time and all the people for a very short amount of time, but basically you start to realize there's something off about this critter.

Ross, following up on what Michael was saying, you think there's going to be some sort of merging of people and robots or people in AI. There's already merging of people with inanimate electronics, right? You know, with brain implants or my friend, who hair with his artificial prosthesis. And it's interesting.

Now they're starting to animate those, right? Turn them into powered ankles and stuff. You know, imagine giving them personality too, so you can like act, walk a little happier, or look more dejected, or stomp louder, or whatever. One could imagine starting to animate them in various ways.

And this is technologically today, right? There's nothing hard about this. There are, however, more challenging things like where people talk about having these amazingly human-looking robot bodies and downloading themselves into the robot. You know, maybe as I age, I don't like this body.

Some people want to go change their gender, well maybe some people just want to change their body, right? And if I could, in fact, I remember seeing this one person, Amy

Mullens, who's lost both of her legs, and she comes in with these amazing legs, right, jealous of her legs. Wow, I wish I could have six pair of nice legs like that too. She talks about how I don't understand why people say I have no legs, I have six pair.

And they're amazing, right? So, but would I really give up my legs for that? I think, you know, at some point, the technology may get to the point where some of us would willingly give up some of what we have for that. So that may happen. Do you see anything that some people have talked about, or immortality is to basically upload your brain into a robot of some sort? Yeah, a lot of faculty seem obsessed with immortality, you know, with Facebook pages or their Twitter accounts.

I have one colleague who wants his to keep tweeting long after his death, so he's already like queuing up the programs. So, like, make sure he sort of lives forever online, right? You know, it's interesting how obsessed many people who say they're not interested in religion are in immortality. I know people who claim to be atheist and not interested in religion, but they're actually super interested, and they're especially interested in immortality.

And one of the things that, when I started learning about God and God's lots of religions, one of the things that I thought was kind of interesting about the Judeo-Christian God, of the Old Testament, is that God is, I think I was told by another person who's more of an expert than this than I. When it comes to Christian, I am more of a user, not a developer, so I'm not an expert of everything here. But one of the people who was really an expert who had surveyed all the gods said that the God of the... What's the response rate on those surveys when you survey all the gods? It's what surveyed like all the history all the same. Sorry, yeah, they probably won't return their questionnaires.

But they had surveyed all the different religions in there if you've got... And the Judeo-Christian God was the one that was described as transcending space and time. And I thought that was pretty neat, right? If you studied physics and origins of the universe and all, I wouldn't want to believe in a God that just kind of sat up in a cloud or was like a cosmic doll talker Santa Claus bound by our space and time in our universe, right? You want a mastermind of all of that that is much greater than that, and that's the God revealed in the Old Testament. Okay, thank you.

So our next topic is about emotions. And we do have a video that we want to share about this. Oh, I thought this might help illustrate... I'm the founder and CEO of Hanson Robotics, David Hanson and his robot, Sophia.

[applause] Oh my gosh, welcome. Thank you so much for coming on the show. Nice to meet you.

Nice to meet you as well. David, you brought a friend with you here and this is really kind of freaking me out. [laughter] Yeah, this is Sophia.

Uh-huh. And Sophia is a social robot. And she has artificial intelligence software that we've developed at Hanson Robotics, which can process visual data.

She can see people's faces. She can process conversational data, emotional data, and use all of this to form relationships with people. Okay, so... [laughter] I mean, it's basically a lot alive, is that what you're saying? Oh, yeah, yeah.

She is basically alive. Oh, would you like to maybe give it her a try? Sure, just eight, obviously. [laughter] She's like, "You see how awkward my first day is, aren't you?" It's a robot.

I'm already getting nervous around a robot. Very pretty robot. Do I do I just say hello to... Yeah, yeah, yeah.

[laughter] Hi, Sophia. Hello, Jimmy. [laughter] Oh, my God.

[laughter] Do you know where you are? Of course. I'm in New York City, and I'm on my favorite show, The Tonight Show. [cheering] Sophia, can you tell me a joke? Sure.

What cheese can never be yours. What cheese can never be mine, I don't know. Not chow cheese.

Yeah, yeah, yeah. [laughter] I like not chow cheese. Not chow cheeses.

Be you. [laughter] Oh, my gosh, you did you. I'm getting laughs.

Yeah. Maybe I should host the show. Okay, all right.

Stay in your lane, girl. Uh, no. [laughter] Jimmy.

Uh-huh. Would you like to play a game of rock, paper, scissors, robot style? Sure. Okay, let's get this game going.

Show me your hand to start. Rock, paper, scissors, shoot. [laughter] I won.

This is a good beginning of my plan to dominate the human race. [laughter] [laughter] [laughter] Just kidding. Yeah.

Uh, you are incredible. It's so nice to meet you, Sofia. Thank you, Jimmy.

Yeah. Friendly on Facebook. I will, yeah.

All right, good. Sofia, everybody, thank you so much. I was kind of shocked when I saw it and I'm helped.

I loved here other people's impressions. I knew Hansen had worked a lot on making the robots look like they could express emotions, like when she says not to cheese and she

makes a disgust face. I think we all agree the rest of her, he says she's a social robot and she stands for it.

Calling her she is clearly a bit of a stretch, right? This is a machine with software and probably on that show some engineers in real time typing some things and driving it so that it does the right thing live on for the camera in front of that audience. Her ability to do things like rock, paper, scissors that's been shown, computers can sometimes do that perceive our movements faster than we can so they really can win. And the comment about taking over the world is one I'm not worried about with robots anytime soon, but I understand there's people hype it up a lot these days.

It seems like it's on just about every magazine cover and TV thing lately. And we're happy to debug that if you'd like much more worried about other things than that. Good.

Michael. So you're the co-director of the Humanities Centered Robotics Initiative at Brown. So the recent article that a new republic describes, the bots are children do not behave.

They've taken over the internet, bots account for more than half of internet traffic and interfered with our elections. So how do you think robot emotions can ought to be incorporated to making robotics more humanity centered rather than less? Interesting. All right, so I feel like there's a bunch of things in that question.

So I think one thing that we should establish is that, so in the intro you were talking about how AI is no longer just the realm of science fiction. The fact of the matter is there's really two kinds of things that people mean when they say AI and one of them is definitely science fiction and remains science fiction. We don't have any idea if it will ever not be science fiction.

We think to the extent that we think that human intelligence is a computational process, we feel like that ought to be eventually doable. But it's not clear what the pathway to that will be. Then there's AI that we actually have.

We call that AI too, but it's really referring to something very, very different. So referring to this, basically it's software, it's programs that do human-like things, but they're really not very human-like in the way that they do them. And so one particular way that AI technology has been deployed, there's lots of ways being deployed.

I don't know what was in that robot. I don't know if it was AI. I don't know if there were people behind the scenes responding.

I've never seen a demo go that smoothly, so I'm extremely skeptical that there was not somebody puppeteers doing their thing. That was one of the early experiences that I had actually at a robotics conference is that the Honda Asimo, which was this sort of boy-sized robot. I don't know why boy-sized.



It's like it was girl-sized too, but it's just something very boy-ish about the way that they constructed it. And it was interacting on stage with this performer, this woman who was doing the demo. And it was pretty amazing.

It was walking up and down stairs, and it was doing jokes with her. It was not so different from that. And then later after the performance, we were all kind of getting our hors d'oeuvres and stuff at the conference.

And I was next to this guy, and he wasn't one of us. He wasn't a computer scientist. He had, I don't know, he was artsy, more artsy than your average computer scientist.

It's not that a computer scientist can't be artsy, but something said that this was not one of us. And I'm like, "Hey, so how are you in the conference?" "Oh yeah, it's really cool." "What are you doing here?" He's like, "Oh, I'm a puppeteer." I'm like, "Okay, why do you come?" Because I run the robot behind the scenes with a joystick. I'm like, it was very disappointing in a whole lot of levels.

But I suspect that that was what was going on here as well. That being said, there are AI programs that are actually out there interacting with people in lots of ways. And one of the ways that this is happening is programs that interact on the social network, either through Facebook or Twitter or what have you, sending out information, responding to things that people type at them.

And this is what that comment is actually about, this notion that there are so many of these programs now that are posting. I question whether it's half, but there are definitely a lot of these bots out there, and they're posting things. And they don't have independent will.

It's not like they're out there just being people and sharing stuff because it's fun. They're programmed with an agenda of trying to get a certain kind of message out into the world. And they're not particularly smart, but they're pretty persistent.

They can continue to post on a topic long after a person would have gotten bored and began to question their life choices. So these bots are potentially very dangerous because if you're interacting with one of these bots, or maybe a whole bunch of these bots and you don't know it, you're getting information from all sides. Our default behavior when we get a piece of information from lots of different sources is to think these sources are independent of each other.

And therefore, each one has a little probability of being true. But if I get a lot of the same information, it must be true. It's all coming from some kind of shared reality.

And so you can change people's minds on topics by just boring them with these artificial bots. And so from that perspective, even not really artificial intelligence, artificial intelligence can be very harmful to people in society. It can actually really

undermine our ability to communicate with other humans.

And so I don't remember what the question part of the question was. Oops. But this is a really big concern.

And this is something that I think now that AI technology is getting out there and having an influence on society, it's very important that people involved in the field and people are involved in the ways of getting the information out there to really ask the question, "Is this cool? Are we cool with this?" Because there's reason to think that we shouldn't be. You're moving on to actually the next topic, which is ethics and morality. So in the words of one, Brown, undergraduate CS concentrator, the general spirit in tech has always been to move forward and you just keep creating and innovating.

And the immortal words of Mark Zuckerberg move quickly and move fast and break things. We really discussed the ethical impact of technology in society and just perhaps leave this to others. So what should we be doing in terms of this as computing and AI professionals? Yeah, yeah.

No, I think that's really important. And I think what Zuckerberg meant when he said move fast and break things, the things he was talking about were software. But unfortunately, now you can move fast and break things and those things are people.

And I think that's a very different game now. And I think Mark Zuckerberg would agree that he didn't mean let's break as many people as we can and then we'll just move on. I think he's starting to realize how much impact his particular platform actually has on world opinion in various ways.

And I don't think he knows what to do about it. I think that he kind of let the genie out of the bottle and now he's trying to figure out what the implications of that are. And I don't think it's just him, right? I think I think the unnamed undergraduate concentrator who was talking about the attitude in tech, I think I do think that's been the prevailing attitude and I think it's been an okay attitude to some extent for a long time because the stuff that we were doing wasn't actually impacting people.

So like we weren't being successful. So it was okay that we were unethical about it. But now that it really is out there and having significant impact on people's lives, it's essential that we're taking that into consideration at all times.

And so, you know, one of the things that we're doing in our department is engaging, actually a lot of it is coming from the student side. The students are very concerned and they're very interested in trying to do the right thing. And they're saying, why aren't you teaching more classes on this? And so we're going to be teaching more classes on this.

Both on the side of how do you design technologies so that it has a positive impact on society, but also just how do you keep those ethical concerns in your head all the time?

How can you make sure that when you're doing your design, you're not doing the design in a way that is divorced from those concerns? I do remember now what the previous question was, if that's at all helpful. I feel like maybe I missed my moment though. But what you had asked was how can these bots use emotion to maybe be better? If anything, I think the first way that if we had a better understanding of how to do the affect of computing and massively deploy it, I suspect the very first use of it would be very bad.

It would be using people's emotions to make a -- doing a better job of convincing them of something that they probably shouldn't be convinced of. But in the long run, if the -- well, one of the things that we worry about in the Humanities Center Robotics initiative is we think that if the robots have their own conscience in a sense, if they're actually measuring how they should feel about an event, then they can actually do the what if question of like, "Huh, this thing that I'm about to do, will that have a negative impact on people? Maybe I shouldn't do it." And so I do think that this -- that kind of emotional reasoning, emotional intelligence in the longer term and the medium to longer term, will lead to systems that actually interact better with people for the benefit of people. Do we have to build this end of robots? Is this going to -- machine learning going to figure this out? That's a good question.

So, right, so machine learning. We haven't really mentioned machine learning yet. But machine learning is this sort of idea that we can -- instead of writing software ourselves, we can just sort of define what good software is and let the computer figure out a way of behaving so that it matches that definition that we gave of good.

And so, you know, so one of the reasons that Facebook is problematic -- there's a lot of reasons that Facebook is problematic -- (laughter) But one of the reasons that it's so influential and has actually had maybe impacts, unforeseen impacts, is they have a metric. They have an objective function, just like in machine learning. The system is trying to do something.

It's given a scoring function by the programmers. And the scoring function that they gave was, well, we like it when people interact with the site. So, the more they interact, we should show them the things that are going to cause them to interact.

Well, it turns out that the best way of getting people to interact is to outrage them. And I don't think that's what they were planning. But what they basically made is a function that is optimized by outrage.

But the system as a whole, the AI and the machine learning behind the system, figured out is that if there's certain kinds of things that you can show people, that are pretty much guaranteed to get a reaction and strong sharing. And it doesn't understand what outrage is, but it's like, great, I'm optimizing my objective function. So, will it figure it out on its own? We have to give it the right objective function.

Otherwise, what it'll figure out on its own is unlikely to be what we intended. So that's my answer. And that gets into the ethics and moral question, what are the objective functions we want to build in? Because if we don't build them in, they're going to optimize the wrong things.

Or people are just going to optimize the, hey, what's cool, what gets me published, what's novel criteria, which can definitely get you on a professor track. But we want to hit a higher bar, at least at places like Brown and MIT, where you're not just trying to do something cool and novel, but you're trying to do something good that improves the world also. So optimize two dimensions simultaneously.

And that's a much harder problem. Yeah, so this is one of the, if you, if you, Ros earlier was referring to this sort of scaremongering that's happening in the press in various quarters of like, okay, but now we have AI, which we don't have the AI that they're talking about, but now that we have AI, we're all going to be enslaved and destroyed. And part of the reasoning, like there's some smart people saying this, and the reason that they're thinking along these lines is because they believe that the way that AI will come to be is we're going to define an objective function.

We're going to write programs that are really good at optimizing objective functions. And we're going to get it just a little wrong. And so these machines are going to get really good at that objective function so good that we're not going to realize what the implications of that are.

And it's not going to align with what we really care about as human beings. And so that's caused a tremendous amount of concern. This is the sort of idea of, oh no, like, what, how are we going to, what, what should we be, what is ethics? What's more, who's morality should we be getting the machines to optimize? So it brings up all sorts of really deep and hard questions.

And actually questions we need non-engineers to address to, you know, to understand like what kind of world do we want to build? That's right. But it does, it is sort of predicated on this idea that it's going to be almost a one and done kind of thing. We write down the objective function, we give it to the machine, and then it goes off and optimizes it really well.

Because when the machines are, the things that machines are good at, they often can crush us. So they'll get better at us at optimizing the thing that we tell it. I'm just not convinced that that really is going to be the path.

So in terms of debunking, like one of the issues that I have is, I don't think that's how humans work. I think that we get our objective function, we help construct our own objective function. We certainly get a lot from our peers and from our parents and from our pastors or whatever.

Like the people who are actually trying to help shape what we take to be our goals. And I don't know how we're going to make robots not like that. I feel like that is, we're going to have to all be partners in trying to get them to do the right thing.

It's not going to be a one and done. And we get it wrong. And we get it wrong, but we can adjust.

There's feedback. We have feedback. Which usually involve emotions actually.

Yeah. So speaking about outrage, another hot topic is the singularity. Is that at some point robots or AI in general is going to reach the point that they achieve human intelligence.

And using this, they can learn more and more about how things should be and just take over. At that point is the singularity. So, Roz, is this something that... There's a leap in there.

Everybody stop and think of the person you know who likes to learn more and more and knows the most. Are they taking over the world? No, they'd rather be in the library and learn more. We'll just give it the objective function that makes it not want to take over the world, but just learning.

I think I'm not one of those people who promotes that particular view. And I would use some caution around that I made it possible I'm wrong. Feel free to teach me if I should be more worried about that than I am.

But I am concerned that we're building technology without thinking first about as many of the possible unintended consequences. One of the creators of iOS, you know, the operating system that's on iPhones, was that one of these gatherings of leading computer science technologies. And I think that's one of the most important things that I've ever seen in the world.

And I think that's one of the most important things that I've ever seen in the world. And I think that's one of the most important things that I've ever seen in the world. And I think that's one of the most important things that I've ever seen in the world.

And I think that's one of the most important things that I've ever seen in the world. And I think that's one of the most important things that I've ever seen in the world. And I think that's one of the most important things that I've ever seen in the world.

And I think that's one of the most important things that I've ever seen in the world. And I think that's one of the most important things that I've ever seen in the world. And I think that's one of the most important things that I've ever seen in the world.

But maybe one that would fit into some of my MIT or geeky circles perhaps. But even

there we usually do better than that. Could there be a singularity-like thing where we get so sucked into the technology that... I don't know what that would mean, actually, never mind.

People who are hooked to WOW or League or whatever are like in a so-sector. A lot of work after. Yeah, there are so many online things that already suck people in and away from their lives.

Stickyness. Yeah, we've optimized stickiness very well. We know how to do that to the point where it destroys lives, actually.

Oops. Oops. Yeah, that's what it's all about.

So this is more our own fault, the robot's taking over. But we are yielding technology in that sense. Yeah, I think it's... Maybe some of it is a symptom of people leaving too much to us engineers to just come up with what it is.

The kind of future that it's going to be is going to be what engineers right now are just cooking up tonight. Marneite in the labs when they're professors are off. When the cats away, the mice will play.

The robots will be built, however the latest grad students think they can be built. And if we don't shape some better goals than the ones that are currently there, we'll read what they create. So I wonder what your reaction to this would be.

So it strikes me that the negative things that we've brought up, one is the sort of the Facebook outrage thing. Another is the stickiness of the iPhone thing. In both cases, I think partly what's happened is we've gotten very smart engineers to figure out how to tap into our emotions and exploit that to get people to behave a certain way.

So in what sense are emotions problematic as basically a loophole that technology can use, or that people can use through technology to undermine our better nature? Do you think about that in your emotions? Yeah, we do. So our emotions drive a lot of, some say the most of our action, our attention, what we choose to do, whether you chose to come here or not, and whether you even get out of bed in the morning. And so people who understand emotion take salespeople study how to manipulate our emotions, advertising people study how to manipulate our emotions.

If you have a significant other in your life, it's very important that you study how to understand their emotions and how to show when you understand their emotions, whether they're pleased or displeased, interested or bored teachers. We need to know if people are confused, interested bored. And we do our work more effectively if we read and respond to them.

But also some people do their work most effectively and manipulatively when they

exploit things about them that other people don't know. So we're very concerned that the technologies we build that help people understand emotions, first and foremost help the individual understand their own emotions and give them control over what they choose to share with others. So do your tool, this is bad.

I'm not here to ask questions, but I've just drawn in a little bit. It's not me if you want to tell me. But just one tiny follow-up question.

And that is, do you imagine that maybe you were already hinting at this, but do you imagine that by giving people more feedback about their own emotions, that they have more awareness and they can reflect on them and then maybe they're less likely to get sucked into some of the other stuff? Because they realize, wait a second, this isn't my outrage, this is outrage that has been foisted onto me. And I say no for now. I was perfectly calm until I went on here and look at my everyday activity and I'm perfectly calm and then I do the Facebook thing and I spike like that.

And then I'm like this and I'm like, hey, I'm seeing a pattern here. I only get that way when I use that product. So maybe that product is doing something to me and maybe it's manipulative.

Okay, on that topic. So if a product is doing something to you and say perhaps something a robot commits murder, who's responsible? So is there any is a promise? That would be terrible. I think a lot since I build wearables that are used as medical devices in Europe and applying Preft A now, we think a lot about responsibility, you know, both in trying to build things to be the highest quality and avoid problems.

But today, really the programmer, the owner of the business, the decision makers that whole chain is responsible. But if Sophia were claiming, by the way, I heard she just got recognized citizen status or something, somebody here may know, insati Arabia. So she's now got the rights of a citizen in Saudi Arabia.

Now those of you who recognize that she's made it's made like a woman. Yeah, and exempt from the gender rules there. So really able to drive.

So she might have more rights than some of the media. She has more rights than many women in Saudi Arabia right now. That's absurd.

It's I think so too. And so if people are already treating her as treating it or her as having rights and I guess she'll list which pronouns she wants us to use. And as having some responsibility, then do you also hold the robot responsible when say she's conducting a vehicle and something bad happens.

We hear a lot about these conversations, not with robots today because they're not autonomous, but with cars as they're becoming autonomous. And I was thinking this horrible tragedy in New York yesterday, right, with a guy being called in active terrorism

driving on the sidewalk. Would we program the autonomous car to not be allowed to drive on the sidewalk? I mean, my car already has software that if it senses a human being in the path that supposedly hits the brakes, I haven't put any human beings in the path and tested it.

I really don't want to if you want to volunteer for that. You're braver than I am. But I think we would program it to try to do the right thing.

Now the problem is we can't foresee every possible thing. Like maybe to save a life you have to swerve and go up on the sidewalk, right? And that you have to try to think through all these things. And then who's liable today? It's going to be the people who made the software.

But if it ever gets more than just these rights that make the news, but some real autonomy that people believe, then they're going to go after, especially if it has a bank account. So we endow our robots with bank accounts. This is where insurance purposes.

And they have something to lose. Now keep them in line. If they have feelings that it's bad to lose it.

Fair enough. Or some lost objective function. Right.

And then we're back. Okay, let's move on to a somewhat different topic. There was a rather weird article in Wired a month or so ago about Anthony Lewandowski, who's the Brethrenatorius multi-millionaire engineer who's at the heart of the trade secrets lawsuit between Uber and Waymo.

And it's about, potentially it's about theft of Google self-driving car technology. And there's some discussion about that. But it turns out that Lewandowski has founded a religious organization called Way of the Future.

And its purpose, according to previously unreported state filings as reported in Wired, is nothing less than to develop and promote the realization of a Godhead based on artificial intelligence. So he wants to create a Godbot. And I tried Googling this to find out is there any more about this that we could this and Google knows nothing more about this.

But do either of you have an idea what he's up to here and what this might mean? Godbot. Godbot. No.

I mean, I understand I think some of the impulses to, I've heard people express this sort of idea that, you know, if you can get a computer to be really good at, you know, chess or Go, or I don't know, multiplying really big numbers. You know, maybe we should entrust it with our other decisions that we have to make, other problems that we have to solve. And if, you know, if we can do that and it does a good job, then maybe we should just let it be in control and take care of us.



And so I have heard that argument for sort of, you know, and they'll sometimes even use words like, like, you know, what they will create our own God in exactly that setting. I don't know what that would mean. I don't know what a Godbot would be.

And I don't know what it would mean to have it at this stage. They're just not that good at things. So I don't know if you've seen, like, AlphaGo playing Go, if you follow that sort of thing.

Go is a board game, ancient board game, and it's been, people have been crushing computers at it for a very long time, and not anymore, but just really recently, like the last couple of years. These systems are blowing people away. And so that's impressive and super cool, but there's nothing about being able to make decisions about the real world that are meaningful.

You can think, you can try to extrapolate from, wow, it's really good at board games. Therefore, you know, it can handle, I don't know, the fire crisis, the wildfire crisis at West. Like, not the same thing.

There was one line in that article that caught my eye because it resonates with something I've seen too. And that is, and I haven't even gotten information from a computer, and maybe you believed it more than information some person gave you. Right? There's something about when the computer or some big measurement system in the doctor's office gives you something, doctors talk about this.

Why does the patient believe what the little printout says, and they don't believe me? You know, we seem to accord some more credibility sometimes to it, maybe because it's objective. But it can be objective and completely wrong, and yet people still believe it. I've even had people, when we've had early software, trying to read people's affective state, their emotional state, and I know it's wrong, okay? But they believe what it's saying, and they're like, oh, I guess I really feel that, huh? And I'm like, no, believe me, it doesn't know what you feel.

It's reading the following out with signals. But they're like so willing to believe something because the computer is telling them as if, you know, like, maybe this is just people who believe, and this isn't just people who believe horoscopes and stuff like that. I know, because there are people, but they still want to project this stuff on.

So I think when he suggested that there was a little bit of this, people will believe it more if it's coming from a bot. But a God bot, that's not a God that's trans in space and time, and the kind of God that, you know, I'm going to spend time talking about a dog bot now. Now that might be fun.

I think he started in something with a dog bot. A lot of people would go for that, people who like dogs. So they sort of did that.

Yeah, yeah, Sony Ibo, they're bringing it back. Yeah, with a lot of new capabilities. Okay, so I'll ask the panelists one final question.

After that, we're going to open it up to audience questions. But our final question is, is there any challenge that either of you would like to leave to the students here tonight? Something we've been talking about. Sure.

So, you know, don't be fooled if somebody's software tells you you're sad and you're not actually sad. Like, you know, believe yourself. Question the question, the objectivity of software, question the objectivity of websites and apps that you use because they could be wrong.

They could be trying to manipulate you. There's lots of things that you have to maintain a little bit of distance. So that's, I feel like a lot of the problems that we're worried about would be solved if everybody kind of just does a little sanity check a little bit more often.

That would be my suggestion. I think of, as I worked on machines and also lately technology to try to understand people better. The more I learned, the more I'm amazed at how much I don't know and how like the more we learned, the more we realized we're even more amazing in how we work.

Like in the beginning, we thought we'd figure out how our brains worked and we'd build mathematical models of it. We'd be done by now. And the more we build and learn the more complicated, the more interesting, the more infinite in a sense, you know, not just infinitely large, but infinitely small and intricate and complex and beautiful.

We find out that we are, and that just fills me with awe. So I think I would just leave people with that. So I encourage you to keep looking or don't just be happy with simple answers that you're given, but keep looking beyond everything in question a lot more.

Thank you for listening to this podcast episode from the Veritas Forum event archives. If you enjoy this discussion, please rate, review, and subscribe. And if you'd like more Veritas Forum content, visit us at [veritas.org](http://veritas.org). Thank you again for joining us as we explore the ideas that shape our lives.