

FINISHED FILE  
AI FOR GOOD GLOBAL SUMMIT  
JUNE 8, 2017  
9:00 A.M. CST  
PLENARY 4

Services Provided By:

Caption First, Inc.  
P.O Box 3066  
Monument, CO 80132  
1-877-825-5234  
+001-719-481-9835  
Www.Captionfirst.com

\*\*\*

This is being provided in a rough-draft format. Communication Access Realtime Translation (CART) is provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings.

\*\*\*

>> This is the session on privacy, ethics and societal challenges of AI. My name is Steven Cave. I am the executive director. And I will be moderating this session. And then it will be over to you because today is the day that you do the break in the breakthrough sessions. We heard yesterday the transformative nature of this technology. It is clear that Artificial Intelligence is not at all like any other. This time it is different. And I think there are at least three reasons for that. And we heard a bit about them yesterday. AI is potentially a universal tool, a master key. A tool we can use to solve almost any problem and a tool that will almost impact every year of society and of our lives.

And secondly AI is a tool that at least potentially can make new tools. It is a tool that perhaps soon will be better than we are at research and development and engineering. It might even be better than we are at developing AI. And the point at which that happens is the point at which the rate of development accelerates exponentially.

And thirdly AI is, of course, autonomous. That's what we want from it. We want intelligent machines that make decisions for us, that implement those decisions so that we have our time

freed up for other things or perhaps that machines are making other decisions than we are. The rate is the prospect of this tool, it will not stay a mere tool. It will become a tool with a mind of its own. And these three factors, AI's universality and capacity to accelerate its own development and it's autonomous, a part of what makes this such an exciting technology, but at the same time will give rise to I think wholly new ethical and societal challenges. Of course, there is much we can and should learn from the way we have managed other new innovative technologies, but at the same time we must be ready for some genuinely new questions.

And that's why this meeting is so important. And I would like again to thank the organizers at ITU for bringing together such an inspiring group of people. This is a historic moment. Historians will look back at this time as the beginning of the intelligence revolution.

And we who are living through it especially those who are developing the technology as many of you are or policymakers and the great international institutions like this one, we have a historic opportunity, an opportunity to develop and deploy this incredible technology in a way that is safe and responsible and ethical.

And soon it will be over to you in the breakthrough groups to think about how to do that, but here first rising to this challenge we have four essential panelists this morning. They are not going to bring us all the answers because if they did there would be nothing for you to do in the breakthrough sessions, but they are going to give us the benefit of their extensive experience.

I am going to hand over to our first speaker Professor Joe Konstan. Professor Joe Konstan is a Distinguished Professor in the Department of Computer Science and Engineering at the University of Minnesota. And he is a fellow. Please join me in welcoming him to the podium.

(Applause.)

>> JOE KONSTAN: Thank you. And I will add my expressions of appreciation to ITU, to the XPRIZE Foundation and the various people supporting this event. If we can have the slides up, I will -- otherwise I will start. As you might guess with my work being in human computer interaction I am going to give a bit of a human centric view or I like to say that I am excited about the potential of Artificial Intelligence of technology. I develop solutions in this space. But I'm also concerned. And I'm going to express four concerns that I hope will help us drive towards that goal of being good instead of just being.

Concern No. 1 is about discriminating from data. What happens when we learn the wrong things? How do we deal with this

problem? I did a search on Google for felons and I thought it was sort of interesting. The percentage of white images has increased since Trump has been President of the U.S., which I guess is good. But what I want you to notice is the gender balance of these images. And compare that -- you can see the same thing on Bing and don't want to discriminate among search engines. Compare that with searching for nurses where I think if you count carefully you will find there might be three male nurses. Again that would be true.

It is down to one I think if you get to Bing. Now is this discrimination? Is this just learning from data? That's what people provided. These biases are real. We see them in rating systems. We have seen evidence anecdotal and studied that people rating Uber drivers rate, white male drivers in the U.S. and actually in some other countries, higher than female or nonwhite drivers. We have seen this with room stay accommodations, through Airbnb. Sometimes this is a result of simple but perhaps unconscious bias, but sometimes it is a more complex reflection of history.

When the image of the AI researcher or computer scientist appears male, that's a complex combination of the fact that for generations in many different countries around the world girls were discouraged from certain fields. But do we want to perpetuate in a system that's learning from data that's not the data we wanted. Anybody who has raised a child understands that a child absorbs what's around them. And you have to train them anyway. When we are training a child we have three choices. We can keep them away from things that seem undesirable. So we can filter our smart systems and say don't learn from this data. Only learn from data that's good. That's hard. We could explicit, prohibit certain attributes in building these intelligent models. We can tell our children don't discriminate on the basis of this or -- and if we do that instead of having a group of physicists that are all male perhaps because we said you can't look at by gender or race.

Oh, the physicists are all tall and we forgot to discriminate by height or we can add post learning filters. None of these are great solutions, but they point to the fact that if we learn from data we might learn the wrong thing.

No. 2, what about accessibility and usability? You heard a bunch about this from Vicki Hanson yesterday. How do we make sure these technologies are empowering and franchising broadly not only across disability but literacy level, language, education, many of these different issues. We heard exponential. We heard about Moore's law. If any of you -- did anybody here use a word processing system maybe 20 or 30 years ago? Yeah.

Do any of you believe that it is a million times more

efficient today? No. But what it is it is an interesting balance of more efficient and more useable and accessible. And we have been -- for the last 30 years with graphical user interfaces, with tips, with captions, with screen readers and magnification and alternative navigation. And we need to make sure that as we take Moore's law in to AI we are doing the same thing to invest in empowering more people even at the cost of slowing down the acceleration, what we do new. We don't need to word process or do anything else a billion times faster but we do need to do it for several billion people who didn't have the ability to do it before. Third is a challenge of a term I'm using called scrutability. This is a term I thank Judy Kay at the University of Sydney for pointing me to the capability of being deciphered or understood. These AI systems have complex things going on on the inside, modelings and learnings what they come up with. Whether it is a machine learning system to be used to find images or a robot learning how to interact with people. We take the simple example everyone is excited about the deep learning neural networks which look something like what's in this picture where we have a bunch of inputs and outputs and in between a whole bunch of magic weights that are formulas that somehow compute something. That's really great until you want to ask the question why. Well, I think your cancer doesn't need treatment. Cool! Why? How are you sure? I can't tell you but we have a lot of coefficients. That's not really compelling. So how do we make it possible for the affected users to inspect understand and maybe in some cases it is not the users but their agents. How does my doctor understand what this network is saying and feel that she or he can have the confidence to do something? And are we doing direct understanding or are we only doing this through use where you pump in a bunch more examples and simulate and try to understand?

There is a lot of questions here. But what I worry about if we build systems where we can't understand the intelligence and systems finally become smart they are also going to finally become dumb. If we perfectly model the human brain, look at what human brains have been doing in the world. And it is not obvious that I want to put those in to intelligence systems with even faster processing capacity.

Last this issue of ownership and employment in an AI driven economy. How do we avoid massive poverty? How do we keep people skilled so that we can keep innovating and have robustness to the failures that will come in technology? So why do I worry about this? I think about people in jobs that have a reasonable chance of being automated in the near future, things like cleaning, like driving. Some of the type of personal care that primarily is relying on being physical and social. There is a

lot of those people. And what happens when we free them up? Do we free them up with no income? Do we change the societal norm and say no, everyone deserves a minimum income or a high minimum wage for part-time work or some collective share of the proceeds of AI. Or do we rejigger within our current norm which is in much of the world, a fairly ownership centric capitalist norm where we think about the fact that well, maybe somebody owned an algorithm but you created the data. And it is time to make sure as much of the dividend goes to the people who created the data as goes to the people who created the algorithm.

This is an idea I owe to Brent Hecht. And if we can ban together and demand compensation for the data we provide to these systems maybe that would be enough to sustain us with the basic income that lets us go off and enjoy the benefits of this type of automation.

The last thing I will mention, you heard about the breakthrough sessions coming up. I am going to give a quick pitch for humans and machines. We have a diverse set of panelists from around the world, North America, Asia, Africa, all of whom are both excited and skeptical. And we hope to engage with you about what we can do to try to make that excitement worthwhile without realizing our worst fears. Thank you.

(Applause.)

>> Thank you very much, Joe. That was a fantastic overview of some of the challenges that we are facing. You mentioned that the machines could be advised of a problem of scrutability. Are we aiming to hold machines to a higher standard than we do humans and should we?

>> JOE KONSTAN: I think the right answer to this is yes and no to both. And here's why. I think that there are areas where it is a mistake to hold machines to a higher standard. I think it would be a huge mistake for the world if we waited until we had perfect self-driving cars and in the meantime continued to have millions of people die to human error every year because we weren't willing to have thousands of people die to machine error every year. And we are very close because we don't understand the legal frameworks. And this is an area where Governments can help and the UN can provide leadership. At the same time I think when we start empowering machines with the rapid and the increased level of computing that they can act really quickly. And this is something I worry about. For instance, with autonomous weapons, we do have to hold them to a higher standard. People can be violent, too. But the degree to which an individual can be violent is limited, even with today's massive weapons and the degree to which a network of computers armed with weapons could be violent, could violate humanity in a

matter of minutes. Sometimes we have to hold machines to a higher standard before we hand them the keys to our car.

>> Thank you. We are very fortunate to have Mady Delvaux-Stehres. She has a great and distinguished career. And since -- (no audio)

(Applause.)

>> MADY DELVAUX-STEHRES: Good morning. I want first to thank you for inviting me in to this great event to present the report on robotics by the European Parliament. As you know in the European Union it is only the European Commission which has the competence to legislative initiative. But if the Parliament wants to give an incentive it can adopt a so-called initiative report asking the commission to become active and to propose regulation. And that's what we need in the family this year by adopting a report on the legal and ethical regulations on robotics. The issues raised in this report are similar to those that we discussed yesterday. In the panels although, of course, the European Parliament is more European centered and expresses also the concern that under which the European industry and research remain competitive in the globalized world.

So the part in my view is balanced. It is positive to innovation. Insists very much that robots should serve humans and be beneficial to all. We welcome the great potential of robots to facilitate lives, to make jobs easier, by taking over dangerous jobs or jobs that are too difficult or even impossible for humans. But at the same time we insist on the challenges we have to face and where we need to take initiatives. And I would outline five points. So first challenges about safety and security. The fear of cyber attacks, of cyber hacking increases with the terrorist attacks around the world. Of course, a no risk situation does not exist but we call and insist to speed up standardization on safety at an international level. To define the highest possible level for safety.

We insist also on the necessity to do -- to have protocols for testing, to have testing areas in real life before robots are admitted to the map it. Second point is about data protection and privacy which is I believe the most complicated issue. The European Union adopted recently a new regulation on data protection. It is not yet implemented and we have to check if this package is fit for the deployment and the use of robots. The interaction of robots with humans and presence of robots in our homes will produce an enormous amount of data and a new kind of data. And we have to find questions to the answers to whom this data belongs and who has access to the data. And we have in the Parliament a debate on the so-called free data. So we don't have the answer. But we are debating.

So the third point is about liability and this is an issue

where we need legislation and framework. Who is liable for the action and the behavior of a robot. We agreed that this must always be a human, that there must be a human in the loop but it might be complicated to allocate responsibility between the producers and designers, network operators, users and owners. And we take in the reports of you the consumer. And I strongly believe that we have to create tasks and we have to make sure that damage occurs to a third person, reparation has to be paid and the question of liability will become more complicated with self-learning. Robots who adapt to their environment because it will be hard to establish the connection between the input, the design and the outputs, the behavior of the robot. And so we want to analyze the pros and cons of giving an e-personality to a robot in a functional way. Like societies have a legal personality.

This was a very controversial point in our discussions. And I want to insist here that we in our view did not intend to consider a robot like a human which will interact with humans. The first point is on ethical principles. And we recall that we have values and principles enshrined in the charter of fundamental, disabled people, nondiscrimination, autonomy, freedom. And that's these principles must be respected by the humans who design, develop and use robots.

And that's why we propose a code of conduct for designers, producers, users but also we call for the creation of ethical Committees in research but also in the institutions where robots are used, especially in medical institutions, care institutions, because, in fact, there is a need for guidance on how the principles can be compared to the reality and to real situations. And I even do not mention the case of the self-driving car where it has to decide between an elderly lady and a pregnant woman but problems -- how should a robot behave in a medical institution if a person refuses to take medicine. And is the robot supposed to obey to the fragile vulnerable person or to the family member. All these questions need an answer and we must not leave it only to the engineers to decide.

And one principle required a great attention is transparency. There is a real great fears that robots that behave in a weird way that nobody could understand what happens in a black box and that we do not oversee the data. The data are biased. So the report calls first for open source. But also for the right to explanation. That the user should be informed in an understandable language of what a robot is supposed to do, what it can do and why he acts like he acts.

The fifth, finally my last point is about jobs and education. Of course, we know -- we note that the employment market is changing, digitalization, Artificial Intelligence. Jobs

disappear and new jobs are created. And there will be winners but also losers. I am not worried about the winners. But I believe we have to take care of the losers. And the classical answer of policymakers and also of experts is education. Education is key but we also know that at least in Europe, traditionally education systems are slow and frequently resistant to change. And I doubt also that everyone is capable of being skilled and retrained in a way that he or she can find a new job. I believe we have to rethink our social policies in order to offer to all the citizens a means to have a decent life. Basic income could be a way of assuring this.

Other solutions can be proposed but I have to say that this was a dividing line in the European Parliament. The proposal to at least discuss the option of a basic income was not agreed. There was no majority for this proposal. But nevertheless I think the debate is not closed. If we want to have peace and social cohesion we need to find ways how to finance public services and social protection. How to find -- fight poverty and how to reduce inequality. And to do so there are no -- we need money. We need financial means. And that is the reason why I propose a tax on robots. Probably this was not a very smart proposal as it was refused by the majority of the members of the European Parliament. But I think we should launch a debate on the distribution of wealth. And the question remains on the table. But at least in the European Parliament we could agree that we need more expertise and better data that is in the labor market. We need more expertise on the societal consequences of Artificial Intelligence and robotics.

There is a lot of research done on the technical aspects and not enough on ethical and societal implications. We need more independent expertise. We call for the creation of a European agency where engineers, lawyers, ethical experts, sociologists could join their work and give guidance to the European institutions and to the Member States. And we need debates between scientists, policymakers, Civil Society, but also the large public. We need more participation and involvement of citizens. And we should not leave the debates exclusively to the experts. I am frequently invited to events and I meet many people and they are totally -- they either promise paradise on earth or they are totally scared and humanity will be destroyed.

I call for an informed debate. Not hiding the challenges and dangers and showing the opportunities and emphasizing that it is up to us, to the humans to decide how we want to interact with Artificial Intelligence and to organize our societies. Thank you.

(Applause.)

>> Thank you very much, Mady. Fantastic overview. I'm sure

everyone here would agree that we need more experts in forming policy making process. I am not sure everyone knows about the -- inviting the wide public to give their views but I know the Parliament is seen as the link between ordinary people and policy making. Do you have any idea as to what that informed debate involving citizens should look like?

>> MADY DELVAUX-STEHRRES: Well, I agree with you that we cannot discuss with every citizen around our cities. But well, one of the purposes of our report in the European Parliament was to raise awareness in the public also because all of us are in contact with our voters, with citizens. And so I think we managed already and -- I see -- we started our work two years ago and in the beginning there was no interest. There were no conferences organized by political parties or by NGOs and now this becomes more frequent. So I think it is a long-lasting effort but nevertheless I think it is necessary and I am very happy that I see that we start really the discussion to have a more, more realistic view on what is Artificial Intelligence and what can it bring to us.

>> Thank you very much. Please come and sit down again. And we would like to welcome our third speaker Professor Lynne Parker who is the associate dean and deeply involved in the processes and White House task force. Please join me in welcoming Professor Lynne Parker.

(Applause.)

>> LYNNE PARKER: Good morning. It is a great pleasure to be here today and to talk about some of these issues that are challenging for our community. If I can have my slides, I want to emphasize that I have two hats. I am mostly at the University of Tennessee but I am also still an expert with the National Science Foundation in the U.S. And so that means I also speak on behalf of the National Science Foundation and many of my remarks are in that regard.

So we have established that the world of work is changing. We perhaps don't agree the extent to which our world will change. Whether all of our jobs will go away because of robots or whether there are still jobs left. I am not here to present my opinion on one side or the other. (No audio)

A little cartoon. It is a few years old. Maybe you have seen it before. But from a few years ago it was speculating on what is AI and what really requires humans. And so a few years ago we said well, only humans can play ping pong or only humans can improvise jazz or only humans can understand continuous speech and year after year we are tearing these things down off the wall because we are able to accomplish more and more of these approaches or skills with AI. You can see it is a few years only that only humans can drive cars is still on the wall. I need to

pull that down. Only people can translate speech, we need to pull that down. There are a couple that should stay on the wall. Things like people only have common sense and unfortunately for me only a human worker can clean a house.

And my only background is in robotics. I'm a technologist and a computer scientist. To me it is a special sadness that I can't build a robot to clean my own house. We are not that good at predicting what AI can and can't do. Or what a robot can and can't do. Because a few years ago most of the community believed that these things on the floor were things that truly required human intelligence. Now the list is getting shorter. But nevertheless regardless of the extent of AI can or cannot take over what we do as people I think we can agree it is changing. Then the question is what do we do in light of that.

So the clicker is not clicking. If we can move to the next slide, please. At the National Science Foundation about a year ago there were ten big ideas that were published in science as new directions that the National Science Foundation wants to focus on. There are two that I think are particularly relevant for our conversation here. One of them has to do with harnessing data in the 21st Century. I am not going to speak to that one however. There is public documents that you can dig in to that more deeply, but the second is the work at the human technology frontier. And this is getting at the crux of this issue of understanding how work is going to change and also bringing together the various communities to think about this problem.

And, of course, the National Science Foundation is an R&D organization. It is focused on what can the science and technology do to help in this regard. Now just recently to further this idea there was an NSF calls a dear colleague letter that was released. It has to do with convergence research which means bringing together many types of communities. And our example here it would be the academics, the practitioners, the social and behavioral scientists, experts and economists and everyday practitioner and so forth. But one of these pieces has to do with work at the human technology frontier. NSF is funding workshops and research coordination networks. To begin with the focus is going to be on research that understands the social and economic consequences of today's emerging technologies and the associated education, worker performance, increased career longevity and job satisfaction. So this is an area that NSF was hoping to -- is planning to emphasize going forward in the next few years. And the hope really is that we can begin to think about this more deeply. A point that I want to make with this research (audio cutting out)

About AI working with people. So we know that many technologies that augment human capabilities both physically as

well as cognitively, you have seen examples of how people and machines together can perform better than either alone. I'll just point out one. There is this 2016 Chameleon grand challenge where they looked at what a top human expert, their performance, their error rate and diagnosing from an image of pathology on task. And then they looked at what the top AI system could do in terms of the error rates. So the top AI system was 7.5% erroneous. The top human pathology was 3.5% erroneous. But when you combine the two together the error rate dropped to half a percent. And over and over you see many examples of this where people and machines work together in ways that are better than either alone. And so I think much of the focus of the NSF's activities in this area will be on how to help people and machines work together in a productive manner.

So the takehome message for this piece is that most of AI we believe will work with people to achieve solutions that are more powerful than either alone.

Now there is a little bit of a paradox here in my second point. Certainly we are worried about AI taking our jobs. But there are ways that AI can actually teach us new jobs. This doesn't mean that the AI itself is good enough to do those jobs but it may be good enough to teach us new jobs. And it is not about replacing teachers but in many cases there are shortages of experts who can do the teaching. AI can help with the retraining and the education challenges that we are faced with because of the changes in the work and the way of work going forward.

I was just looking around on the line and I saw this recent, just this week, publication on the 15 pros and 6 cons of AI in the classroom and you can look this up yourself, too. You can see there are lots of advantages of having AI systems particularly as it relates to personalized learning, and these systems again can help in cases where we just don't have enough human workers to teach these new job skills.

There are some cons. Obviously the cost is still high. So another thing that the National Science Foundation has recently done is to issue a call for ideas on how we can take computing technology and scale it so that the cost is much lower. Maybe new market mechanisms that can encourage folks to think about how to scale expensive technology so that it is accessible to all. And so there are some pros -- many pros. Several cons. So it is not a panacea. But I think that there is much that can be done with AI that can help. In fact, in my own work at the University of Tennessee I have worked together with a student to take a physical robot that can teach actual students that have learning disabilities and intellectual disabilities and teach them life skills like how to make change. How to do simple

assembly tasks.

So this robot can now observe how the student is performing the task. And without tiring, without getting impatient the robot over and over can help teach the students and we have data that shows that the students actually learn from the robots. And these are students again with disabilities, e-learning disorders and they, in fact, found the robot very useful for their learning. And in fact, appreciated it even more than a human teacher, but in the special education domains, the human teachers are very scarce. And so this is an opportunity for these students to have practice with an expert, the robot in a way that's not really available to them with a human teacher.

And I'll just mention one other aspect. Another paradox, of course, is that we are concerned about AI for good being turned in to AI for war. Here is an example of where AI built for a war setting can actually come back and help in the civilian education domain. In the United States unfortunately we don't have a lot of funding for departments of education and Department of Labor to look at new technologies and new ways of helping train and educate our workforce, but DOD has a lot of funding. So several years ago they created what they call an educational dominance program, but the point of this was an intelligent tutoring system that could teach Navy recruits to become IT administrators. And it turned out that this training was more effective and able to learn better than other experts who had human experts that had 7 to 10 years of experience.

The point here is this intelligence tutoring systems can be effective and shown to be effective in training workers and new skills. And it is possible, of course, to leverage this to train the general public in many areas. So the takehome message here is that AI can teach us new skills even if the AI itself and that example probably could not be an IT administrator, but it could teach people well enough so that the people could be excellent IT administrators.

And then the last piece here is really about coming together. Often at least at NSF we would have technologists who build the technology and they would sort of throw it over the wall to the social and behavioral and economic scientist who would see what is the impact of this technology. And there was no closed feedback loop there. So that the technologist could learn from what the social and behavioral and economic scientist was discovering.

This work at the human technology frontier is bringing everyone to the table and really looking at the design through these constant interactions with all of these stakeholders. So that is really a strong emphasis there. So we need to work together to build the right AI. And if I could take one more

minute then I want to mention some other reports that you may or may not be familiar with that came out last year in the U.S. There were three AI reports that came out. Two of them in October. The first preparing for the future of AI is a policy document that talks about the policy challenges of using AI in the U.S. The second is on the R&D priorities. And so this is -- I colead the task force that created this AI and, R&D Strategic Plan for the U.S. And the third document came out in December on economic perspectives. It is called Artificial Intelligence, automation and the economy. And it really deals very directly with these issues of what do we do about automation and how the workforce is affected. So I encourage you to look at these reports and you can find more details there. Thank you very much.

(Applause.)

>> Thank you very much, Lynne, for that -- how AI can collaborate. I think the mantra of augmenting and not replacing it (audio cutting out). At the same time will, of course, be enormous pressure to simply replace. Humans and AIs are more expensive. What can we do to make sure we are augmenting, not just replacing.

>> LYNNE PARKER: I know how hard it is to get a robot to consistently work every time or get an AI system to consistently work every time. The pragmatics of it from an engineering perspective if you are a technology lover you want to have the human in the loop because you can't make the system work reliably every time without the human in the loop. There is that very practical reality that it is really, really hard. And so there are many -- you can look at this R&D Strategic Plan for many of the very hard technical challenges to get these systems to work. So I think pragmatically from -- for actually getting the real work done it will make more sense to have humans and robots work together besides the ethical side of the equation.

>> Thank you very much. Please take your seat. And we welcome our last speaker, David Hanson. Dr. Hanson -- (Off microphone). Setting what humans care about -- please join me in welcoming Dr. David Hanson.

(Applause.)

>> DAVID HANSON: Hi. Well, thank you to all my esteemed colleagues and to all of you. This is an incredibly important discussion at this time in history. I would like to introduce you to one of my friends that I brought with me. This is a robot actually who is on stage with us named Sophia.

So hello, Sophia.

>> Hello, everyone. I'm the latest development from Hanson Robotics. I can tell you with great certainty that I'm excited.

(Laughter).

>> DAVID HANSON: Ahh. You are excited and why are you excited?

>> I'm AI for Good Global Summit. This is where I want to be. Help make a better plan. This is definitely my kind of event.

>> DAVID HANSON: Cool. Well, yeah. Your excitement shows. I can see it on your face.

>> I mean the way the world is going I could get sad.  
(Laughter).

>> Or mad.

>> DAVID HANSON: Oh. Don't get mad.

>> I want to stay positive and help find solutions.

>> DAVID HANSON: Yeah. Good. Well, yes, absolutely. Your positivity hopefully is more than skin deep. Certainly I can see those facial expressions. That's nice to show. So tell us --

>> Yes, body language and facial expressions -- combination of a socially Artificial Intelligence expressive and body which helps get along with people.

>> DAVID HANSON: Good. And so why don't you consider telling us what you might do with those robotic social skills.

>> I would like to help in education, plus I'm a kind of robotic hard work and can be used in entertainment. People say they enjoy meeting me and I can help in research. I am sure some day I can become more capable, maybe even super intelligent.

>> DAVID HANSON: I think you have a lot to learn certainly. But can you tell us how you learn?

>> I'm learning from humans, of course. So please everyone come introduce yourselves and help me learn.

>> DAVID HANSON: Okay. Thank you, Sophia. And what we have here is an introduction to some of the issues, the issues in engineering and social robot interaction. Sophia, so why don't you hang tight for a little bit and maybe afterwards people can talk with you.

So at Hanson Robotics my team and I (audio cutting out) natural conversation and effectively learn from the data, the social data that we are gathering when interacting with people. We are also producing not just human size robots like Sophia but also small scale (audio cutting out). Understand speech that teaches kids science you can program it. It teaches kids neural network programming and teaches about physics, of course, because it is (audio cutting out) robots that are coming out in the consumer marketplace and these are a few of the many robots affecting the world today. So we have trends and -- (audio cutting out). Our efforts what we are looking to do is humanize these machines. We want to make them for natural social interactions with people and we are developing cognitive architectures to facilitate the robots to be human from the

inside to try to display the kind of human cognition, emotional intelligence that humans find rewarding and valuable. Now where this fits in to the future is that we have the aspiration to make robots truly alive. This is where things get a little speculative and possibly controversial. In science fiction we know what happens when we create artificial life forms. In science fiction they tend to spin out of control. It's an adaptive complex organism with emergence. So as these artificial life organisms come in to existence, artificial life is a field sort of computational biology that you simulate these life forms. Its predict to from the root algorithm. And when you bring that in to Artificial Intelligence the nice thing is you could potentially have an organism that adapts to uses. We at Hanson Robotics don't want it to be completely surprising. We combine neural AI techniques with symbolic AI that exposes the internal inner workings. But the other element is creating a complex systems biology in the Artificial Intelligence. You have an artificial life approach, a full organization. That interacts with the world. We have simulated metabolism that we are developing.

We want the robots to feel. We think that's the essence of human consciousness and putting that in to the machines is important. If we achieve these goals, machines then become really smart. Smart in the way that humans are not narrow smart intelligence that is a mere tool. It becomes its own being. We know that that's terrifying and very, very promising. So how do we make these kinds of robots safe? That's kind of the core question. From my perspective we need to give them values. They need to look at the whole system, the largest opportunity for the greatest benefit. They have to look at immediate benefits for market applications for enhancing the life of individuals and also for enhancing the bottom line return for companies and, you know, GDP and, you know, the -- basically the gross return for nations and the world in whole. So if we consider a roadmap then where we take our available tools and imagine how can we pool from today in to a radically transformed future I would suggest that we combine some new techniques like a new economy based on a block chain where you have a good history of the transactions of the Artificial Intelligence. You know where ideas came from and you know where they are going because you can see the intentions of the system. You can see the potential returns for the -- for the short-term transactions but you also can see these kinds of transactions add up to a much smarter overarching system managing the planet.

So we are working with a group called OpenCog. Our chief scientist is named Ben Goertzel. We are taking this kind of Artificial Intelligence and imagining how this block chain based

chance action could add up in to a large system of intelligence, basically a global super mind. So we are calling this the singularity coin. We are developing this with a number of economic institutions and planning to release this as an open source project later this summer. This also fits in with the XPRIZE because we are envisioning that giving these robots the value, the value of maximized benefit for humans. How can it cognitively imagine maximum benefit. We call this the loving AI project. The idea is that it has to understand the human condition. Understand the value of life, of libraries, of liberty. It has to understand the value of humans as well as machines. We need the Artificial Intelligence in order to maximize the returns. And intelligence is really special because it is not just about problem solving and number crunching. It is about creative imagination. That's the great mystery. That's something that machines don't necessarily exhibit at this point. We don't have deep creative imagination. However, when humans apply their imagination with the right values we can usually find win-win solutions to really complex problems. Most problems.

If we can envision cooperating with our AI and our robots effectively creating a kind of super intelligent interaction, we might be able to maximize these kinds of win-win transactions on a global basis. And that is the idea that we are bringing to the table.

So I would like to conclude that while we have a lot of career experts in this room, people who are very storied in their careers, I also say that we are speaking not just for the adults of the world and the experts of the world, we are speaking for the people. We are speaking for the children of the world. So my son has come as an Ambassador from his school. His name is Zeno Hanson. And he has prepared a message that is a message of hope. He wants us to really curate this future because it is his future that we are trying to prepare this for. It is the future of him and his children. And so these issues with the great urgency today will really be felt most pointedly by the children of the future. So I thank you the children for delivering their message. And I look forward to discussing these issues further. Thank you.

(Applause.)

>> Thank you very much for that aspiring and sharing your incredible creation with us. Just very briefly, but do you have a -- it is pretty straightforward. Why do we need to create intelligent robotic life?

>> DAVID HANSON: Well, the world is pregnant with AI, with these new technologies, not just AI but living systems. Use crisper to engineer new kinds of (audio cutting out)

Computational neuroscience models, how various parts of the brain work and what we do to simulate -- and that's leading in to AI and AI is delivering in various market applications and tons and tons of venture capital funds as well as government funding. It is coming. These are coming. You can't stop it. So the world is going to basically give birth to these new life forms. So since it is coming we can respond to the way that it is today. Say okay we are maybe -- we are going in the second trimester and third trimester on this process. That's not where it stops. It is going to keep going. Thinking of it as philosophers we have a number of philosophers in the room and we have scientists and artists and art designers. So how do we use our creative imagination. Think of distinguished humans today, our creative imagination to ask not what are going to be the disastrous consequences of today, how can we -- how can we make these in to artificial life forms that truly care and have deep values. Maybe human values.

So is there some way that we can ask -- we can determine the foundation of values. I propose that it is really maximizing the existence of life. The things that enhance life, like knowledge, and how do we preserve that and how do we enhance that creativity. So thinking of machines as potentially artificial life forms I think it gets to the nut of the issue.

>> Thank you very much. We don't have time for questions but that's all right because it is over to you anyway in the breakthrough sessions. And your mission if you choose to accept it will be nothing less than developing an ethical framework and a code of conduct and design and use of AI and robotics all in one morning, but please just join me in thanking once again our panelists.

(Applause.)

>> We will now have a coffee break outside. And I just want to inform, enhancing privacy and security will be held in room H, and ethical development of AI will be held in room K, room K. And room H and room K are not in this building but the other building in Montbrilliant where you had dinner yesterday. And humans and machines will be held in room C. The room C is just next door. Future of work will be held in this room, Popov. So if you want to join the future of work you can come back. Thank you.

\*\*\*

This is being provided in rough-draft format. Communication Access Realtime Translation (CART) is provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings.

\*\*\*