

RAW FILE  
AI FOR GOOD GLOBAL SUMMIT  
JUNE 7, 2017  
2:00 P.M. CET

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.

\*\*\*

>> Ladies and Gentlemen please take your seats. We would like to start. Thank you so much. Being myself Swiss I'm compulsory about keeping time. So please take your seats. Let's get us started. So good afternoon and welcome to this Plenary session No. 2. My name is Urs Gasser I am the executive director of Berkman Klein center for internet and society at Harvard University and we launched an AI governance in with the MIT lab. Thank you to all the speakers from the morning. We have a fantastic lineup of speakers. We'll roughly follow the mode of this morning's session, that is we will have a series of short presentations followed by one question that I will ask and if time permits we will then also have a bit of a discussion among the panelists towards the end. So this session is actually a sort of a deep dive on some of the themes that were introduced earlier today, in particular in the session before lunch. And the question before us is now what's at the horizon. What are kind of the next generation of AI based technologies that are developed in the labs but that are also increasingly adopted and people are using. And what do we expect from these emerging technologies. How will the world look like in maybe 20, 30 years from now realistically speaking. I know from the presentations

that I have seen that we are very much following the tone of the second panel to take a realistic look at what is happening right now and probably a bit less buzz words and type. So we'll take kind of a fresh perspective on that topic. We'll not only look at what's ahead, what's the trajectory of the technologies. We will also talk about the impact of those technologies. What are some of the main challenges and again some of the core themes that were introduced earlier today, what are some of the opportunities and how can we harness these benefits of AI for the social good. We'll do so, we will address these questions and issues from multiple perspectives. Again also as a follow-up on this morning's conversation. Acknowledging the need that we have to bring the technologies together with the social scientists, with the ethicists with people who think deeply about Human Rights, about ethics, about the law. And so we will have a very multi-disciplinary session then. Our first speakers, I will briefly introduce the speakers and you have links to the bios and is professor Francesca Rossi who is a distinguished research skinist at IBM Watson research center and Professor of Computer science at university of Padova. She is not only developing AI technologies but thinking about the ethical questions in the development and related to the behavior of AI systems. She has many affiliations, among them is that she is a fellow at the European Association for AI and she also sits on the board of the partnership on AI which was introduced already this morning.

She will talk about the opportunities on the horizon and the promise of an symbiotic partnership between the machine and humans. Francesca.

>> FRANCESCA ROSSI: Thank the organizers for inviting me here to give a talk. I am happy to be with all of you. It is very multi-disciplinary environment. Can you hear me? Thanks. So I think that's the only way to take -- talk about these issues in a very multi-disciplinary environment with moving forward one initiative, one event at a time. And being concrete about the next steps.

So yes, so I will talk briefly about the opportunities I see on the horizon but also the challenges, especially the ones related to ethics. So the main point is that I think that the future but even a little bit present but more in the future will be about some symbiotic partnership between humans and machines and the reason why think that is that more and more in academia, which is my whole career has been academia but now in a corporate environment, I see this is the best way to achieve the best results in terms of using AI in our every day personal and professional life. So we have seen a lot of transformations since the industrial revolution. You -- everyone knows about

them, we have seen electricity coming and automobiles, mass production and then information telecommunication and now we have seen this great revolution which I believe is possibly greater than all the previous ones, which really we are -- we have the opportunity to greatly augment human intelligence. With machines that based on the AI techniques can really learn over time how to improve their behavior. Can reason in ways that we humans have less capability to reason. And also hopefully and necessarily they will interact with us in a very natural way because without that natural interaction then the symbiotic relationship cannot take place. Why I do think the future is human plus machine together? And not just machines alone? Because I see more and more in many deployed AI systems that this is really achieving great results. In health care, in various kinds of industries, in education, in space exploration and in many other application areas. Ant main point, the main reason is that I really believe humans and machines are very complimentary. So we are very good at asking ourselves some interesting questions, machines are less coded. We are very good at common sense reasoning, which is understanding how the world functions and reasoning on top of that understanding. We are also very good at value judgment, based on our also emotional situations. But on the other hand we are less good at other things that machines are much better at than us. So machines are much better at patent discoveries, statistical reasoning, large scale mathematical reasoning and so on. So they can really help us in those aspects that we are not very good at. So that's why this complementarity between human form of intelligence and the machine form of intelligence is what has to be exploited in order to get the best result and in order to make better decisions. Our life is full of opportunities and needs to make decisions in everything that we do. And so it really important to be able to make better decision. And what do I mean by better? Better depending on the context and the environment can be more efficient. More confident because based on more knowledge coming from more data, more grounded, more insightful, making decisions that can consume less resources. If you think, for example, at our planet or even in very less scale environment. And even more ethical decision and less biased. So better may mean many different things in different contexts but I think that these synergy between the human and machine form of intelligence will really bring all of that. I think it has the potential to bring in that. So let me give you a few examples that are existing already. So here, for example, is an example where Watson and IBM is helping doctors to find the best diagnosis out of some analysis of -- in this case it is breast cancer kind of diagnosis. So, for example, here the doctor can

conduct the diagnosis and can decide whether it is an lipoma or necrosis or whatever are the options. And then help by the machine the doctor, double check whether this is the most plausible thing and if the machine thinks it is not then can ask for an explanation and give feedback on why -- why some other diagnosis is actually the best explanation.

So here the doctor is really helped by the machine in trying to understand what's the best diagnosis in this particular case. But also learn more general concept that can help him in the future as well.

The second example is about ALS. And so in this case in collaboration with doctors, starting this dangerous disease, the doctors have examined some genes to understand whether they are responsible and correlated to the ALS. But they -- they can't examine all the genes. So they asked the machine to rank the genes according to what the machine thought and they look at the top ten ranked Genes. And eight out of top ten are found to be correlated with ALS and five of eight are not known to be correlated with ALS. Here the machine is helping doctors discovering new correlations between genes. And the disease and so it helping them to focus their reasoning because otherwise the doctor would have to, you know, go through all the genes or randomly look at them without having some initial ranking that the machine based on all the data and information can give for them.

The third example is about creativity. Machines cannot only help in checking data, to help doctors but also checking, you know, the creative landscape to help artists to develop and to create. And so in this case this was in Barcelona where the machine looked at all the works of Gaudi in Barcelona. Look at the style of Gaudi. And then suggested according to this tile, some shapes and forms and materials to use to create something and this is what the artists came out with based on this suggestion and interaction with the machine.

So another example which is more general is that IBM we are also focusing on social good specifically which is mostly related to this event here. So we have a program which is very interesting where students and post-graduates from various Universities come to IBM are mentored by inside researchers in order to tackle a project to develop a project related to some social good initiative. So in collaboration with some Institute, some Foundation organization that can support with the data they tackle one problem and I give you one example. For example, here about the Zika virus. This was a project that was done in 2015 where a student using machine learning and data science together with IBM researchers they focused on identifying primate species that can host the Zika virus and this is helpful to understand

the strain of the virus and how to cure it better.

So with all these potential, of course, we know that whenever we have such a powerful technology there are also challenges. And as we mentioned in the morning already these challenges can be of many forms. Some of them are technological challenges. So one of them as we mentioned algorithmic accountability or capability for these algorithms based on AI techniques to explain why they are suggesting a certain diagnosis, for example, and not another one. Bought because without that capability of explaining to the doctor why they are doing or not doing something or suggesting or not suggesting something then the doctor will not be able to trust the machine correctly. Of course, we don't want the doctor to trust no matter what the machine says. We want the doctor to achieve the right level of trust in this machine. So the machine should be able to explain why it is suggesting some diagnosis and not another one. So that the doctor can know the potential of the machine and also the limitations. So not necessarily everything that comes out of the machine is good because as Yoshua said we know especially techniques based on machine learning they are very good in various tasks and they have a very small percentage of error but they always have some error and the error that they have not what you would expect from a human being. It is a very strange kind of error. Like the two pictures we saw this morning of things that look similar to us but they were interpreted differently by a machine. So definitely algorithmic accountability and strainability is very important and we still are not, you know, are not there yet to understand how to make them in the best way. And then also to make sure that when we deploy an AI we have to make sure that they works according to values that are lined to human ones. If want to give a doctor a decision support system that will support the doctor in whatever decision is to make, diagnosis, therapies and so on and you expect the doctor to follow some ethical guidelines, you also expect that machine that decisions super system to be aware of those ethical guidelines and to follow them while making those suggestions. You need to understand how to embed these ethical values in to a machine. The data bias problem has been mentioned. Because data is everywhere. But -- and the more data the more the machine, you know, as behaves with less error but it doesn't mean that it is not biassed. It doesn't mean that it is diverse enough. That it has all the features that you need to capture really that space of the decision making. And so this ability to be aware of this biases, possibly mitigate them and alert the final human decision maker that there could be some bias introduced by the data. It is crucial in making whatever decision capabilities provided by the machine then it is taken

with the right level of trust and not more or not less. And there are also some global responsibilities that have to do with the ethical development and the deployment of the AI system and impact on the workforce. There are various things that we can do as companies or even noncorporate environment. We can define principles and guidelines and recommendations on how AI should be developed, what the purpose should be. For example, for IBM the purpose is that we don't want to replace human intelligence but augment it and trustworthy of how we deal with the data of humans used to train the machine learning systems and we want to support, you know, this evolution in the workforce. But in general I think that devise many principles and in the last two years here is a complete picture showing all many initiatives that have started in the last only two years or even less that have all very overlapping goals. So understand the impact of AI on the society on people, on organizations, on Government, on policy making, and so on. And so understand what it means to ethically develop, deploy AI and how to embed ethical principles in AI system. Value alignment, data issues, data privacy, policies and so on. And all of them are very related to each other. And in fact, one of my goals in the short term is to try to understand the relationship between all of them because we don't want, of course, this initiative to do redundant work but want to be complimentary and join forces and -- but this is just to tell you really the whole world from different perspective is tackling these issues that we are discussing today. I will speak just more in depth about the partnership on AI that has been mentioned already. So I represent one of the founding members there, IBM. But you may know that this has been funded by six companies which are the six that you see in the top row here, Amazon, Apple, Google and IBM and Microsoft. There are many other partners. Some of them are companies but also many of are NGOs, Civil Societies academic Associations and so on. They want to study impact of AI and educate people, policymakers, government Associations about the real potential of AI, the current state of the art, the limitations, what policies can or should or should not do about regulating AI research, development and deployment. And in general and this has to be done in a very multi-disciplinary environment. However the role of the companies here is to provide the voice of the customer. So the voice of the real life deployed AI systems that once you deploy them in to the real life you really can see the impact and you really can see the challenges and the issues that are raised. So that's where the companies can play a role here and then they can discuss together with everyone else. So to summarize, for me the future is really about, you know, human plus machines. I have seen many, many examples where this is the

way to go if you want to have better decisions. Whatever better means to you. And it is not the human alone. It is not the machine alone, but the human and the machine together and I think this has the best potential to really solve the societal challenges that we face everywhere in the world to close or mitigate the digital divide and also to be very impactful in Developing Countries. Because whatever I mention this health care domain application that I mentioned in my talk I always think and the examples that I have are from our society, like first world society and you can see some improvement there. Helping doctors to make better decision in this kind of health care system. But, of course, the potential the real potential is to help doctors in the Developing Countries where they don't have the support. They don't have the kind of information that our doctors have. So that's where the real impact will be I think. Okay. I'll stop here for now. Thank you.

(Applause.)

>> URS GASSER: Thank you very much, Francesca. If you don't mind to grab a microphone I have a question for you. I liked very much your framing to think about humans and the machine as some sort of a partnership. And partnerships, of course, involve some sort of bidirectional conversations and interactions. Now today naturally as you focus on AI much focus is on the technology side of this partnership, right? It is about the machine and we learn where we stand, what the future brings and the like. But I was wondering from your perspective, particularly being a computer scientist, what somehow the human dimension? Where do you see things going in the future? What do humans, what do we as individuals have to bring to the machine to make it a productive relationship? What does it mean for our mindset for how we think about this interaction?

>> FRANCESCA ROSSI: Yeah. So when I presented this idea of the partnership between human and machine, I described some examples of things that we can provide. We can do much better than machines and vice versa. So, for example, asking the right questions, this is something that we do very well. Or even, you know, value judgment or common sense reasoning. These are still things that we don't know how to embed in machines that I don't see a very short-term solution or how to imbed these things in to machines. While -- so these are the things that we really can provide. And without them the machines will not be able to make good decisions by themselves. And so really we are needed and we are needed to the machines with these things that we can provide in order for them human plus machine overall system to make a better decision. And also this will, you know, in some sense free us from, you know, trying to do the other things that machines can do naturally much better than us, like, you know,

very complex, you know, repetitive reasoning or statistical reasoning. Our brain is not that good at doing that. And so we can focus on the other things that we are very good at that we can provide.

>> URS GASSER: A risk that somehow we start actually to Delegate our questions and our judgment to the machine. Because we ask Siri if we look an answer. So yes, we ask questions but somehow the judgment is increasingly Delegated to autonomous systems. Do we need some sort of a new or enhanced skills of critical thinking as we interact in this partnership? Would you say there is also a kind of a mental process required that we need to readjust looking at some of the examples that you provided?

>> FRANCESCA ROSSI: Of course, there are many things this we are Delegating to machines. Using a calculator to do complex division that before people were doing by hand and now we don't know how to do it. It requires some more time. So we are already Delegating things to machines. But I don't think that we are very close to Delegating this really very human specific, you know, capabilities.

And overall I think that this are -- there is a need to evaluate the overall hybrid, you know, human plus machine system rather than evaluating the error of or the behavior of decision making capabilities of the machine alone. We should never think of the machine alone. But the two together which brings in to place also these capability of interacting of building trust between the human and the machine over repeated interaction over time.

>> URS GASSER: So Francesca's presentation also highlighted some of the ethical challenges and questions. And I think that's a terrific segway in to the next presentation. And really delighted to invite Salil Shetty to join the podium. He will talk about the idea of a new ethics for AI that is rooted in Human Rights and Human Rights principles. So there is a really strong link between these two presentations. So Salil is the Secretary-General of the other AI, the Amnesty International movement where he leads the work to end Human Rights violations in many parts of the world. And that's another connection point. He has done tremendous work to bring Human Rights agendas and issues at the forefront also in the global south before joining AI he was a director of the UN millennium campaign and was deeply involved in the promotion of the Millennium Development Goals. Salil without further ado.

(Applause.)

>> SALIL SHETTY: Thank you. Thank you all for giving me this opportunity. I must be one of the few sort of interloopors in this audience of technologists and policy experts on

technology. I was going to mention, the post lunch session is always the most difficult one. So thank you for bringing us in on this one. But Francesca's talk has certainly woken me up. That's the worst place to be sitting. Because the sofas you start reclining. Do keep an eye on these people to see if they are an I wake. One of the biggest problems we in Amnesty International have with AI we have the same name. I am often confused when people talk about AI inside of amnesty as to which AI they are referring to. We have two AIs, Amnesty International and Artificial Intelligence. You have heard and you will hear a lot more in the course of these days about the amazing things that Artificial Intelligence could do in the future.

I want to talk to you more about Human Rights today and tomorrow. So no surprise on that front. And I want to talk to you about this in a kind of setting which makes it clear that as far as amnesty is concerned we are very clear that there is huge possibilities and benefits from Artificial Intelligence. So the AI for good is an amazing idea of having this conference. The question is who is it good for and we are going to talk more about it in the coming period. I think in the eyes of the general public and certainly within Amnesty International as well it is difficult for people to connect what we do as kind of bread and butter work. Our bread and butter work as you can imagine is freedom of expression stopping torture, that's the sort of work that amnesty typically does. But we are looking more and more equally at the issue of Human Rights in the workplace, access to health care and education, economic and social cultural rights as well.

And in that context I want to just refer to one anecdote and I want to come back to that at the end of my speech and this is I think two years ago I was in a rural part of the Czech Republic and I was taken to a school to look at the situation of Roma girls' education and this is an Roma family and I spent some time talking to this girl. She must have been 10 or 11 years old and I think her name was Aka. You could see from the eyes how bright and how enthusiastic she was. But at the same time the story that she had to go through A to get in to that school as an Roma girl to get in to a school in the Czech Republic you can imagine this is part of the European Union is not the easiest thing to do and as soon as a group of Roma children had managed to get enrolled in to the school, a lot of the non-Roma children left because they don't want to be in the same school as the Roma children. So let's keep Dinka there and we will come back to her later. I do believe that the discussion on Artificial Intelligence is at a fork in the road and I think we have clear choices in front of us. And that we want to be discussing these choices and one of the things which Marc

Mc-Ganda had talked about from our country was the concept until there and until there concept whenever you want to make a decision the way in which you could make it is by thinking about that last person as to what does that last person -- how does it affect that very last person. Who is at the end of the row and in some ways Dinka the Roma girl or a girl in Afghanistan that's a good kind of check in to see as to what does it mean to this girl. So humanity may live on one planet. At least for the time being. But we inhabit different worlds. At one end is prosperity and wealth, technology that's continuously evolving to fulfill any need we can imagine and the future prospects of incredibly long and healthy lifespan space travel and much more. At the other end is the reality of poverty and injustice that has scarcely improve despite the arrival of the Internet and mobile and Big Data and now Artificial Intelligence and in between them is one of the biggest threats of health of our society and that's inequality. If we look at global development indicators there is a very positive picture. Extreme poverty, undernourishment, access to education, clean water is increasing. The world is achieved significant progress in the past 25 years. But at the same time economic inequality has been rising and I always also talk -- when I talk about economic inequality I always try to refer also to voice inequality that those who are left behind have no voice. Even in the OECD countries including of the wealthiest countries in the world, income inequality is at the highest level in ten years. Nine times of that poorest 10%. Up seven times from 25 years ago. These numbers have been thrown at many of you. The one which came out earlier this year, that eight individuals, the wealth of these eight individuals is more than half the population of the world which with 3.6 billion people have as much wealth as eight individuals. Around the world from Asia to America and from Europe to Middle East we are seeing what happens when inequality grows and it is ignored. Anger simmers and the politics of blame thrive, discrimination against women, Muslims, black people, states, dividing the world. Inequality -- when we need to work together to solve the global problems we face but growing inequality is neither an accident nor is it without consequence. Income inequality and political instability generates direct consequences of human choices, Government policy and corporate practice. The way companies and Governments handle technological innovation is a key part of this story. To take one example, one of the most important if he none men none over the past few years, 1.3 people working in the gig economy and they have no guaranteed working hours or sick pay. They lack many of the rights that workers in regular jobs enjoy and at the same time technology companies have made

billions by providing platforms for those so-called flexible working arrangements. Of course, businesses can and should innovate but are they paying their fair share? If you look at growing corporate profits together with wide ranging cuts to health care, education and other public services even in the west, we know that something isn't working. Tax evasion tactics have allowed hugely profitable companies from paying very little tax while whole job categories disappear and are replaced by low paid insecure jobs. This is a rule in life. If you take and take and don't give back people will get pissed off. Some will argue that for business what matters is growth and profits. And this has to be their priority. But then they can't complain of the growing disillusionment and reducing trust levels in business and governments. Societies must plan for and mitigate against the risk and this is critical in the case of Artificial Intelligence which has the potential to create huge economic disruption. As economic inequality grows there is another kind of technology fuelled inequality that's growing at exponential rate that relates to data. There is an already a huge *asem mi tri* in power between companies and Governments and individuals and Civil Society on the other. Control a handful of companies and Governments exercise over unimaginable amounts of personal data. Whether it is to sell as ads for electronic surveillance programs data gives formidable power to those who control it and we heard how Big Data analysis and micro targeting of waters was used in U.S. elections and the Brexit referendum. While the use of data in politics is not new the difference is how powerful these techniques have become in a short space of time and how Artificial Intelligence could super charge all of this. So what will happen tomorrow? Let's fast forward 20 years. What world do we want in 2037? Technology is chief among them, Artificial Intelligence will shape tomorrow's world. That's a certainty. And I don't want to present a binary picture but let's for the purpose of this conversation visualize two scenarios. If we continue as we, hundreds of millions of jobs can be lost to automation and largely replaced with insecure jobs with little protection of workers' rights. We may have some social protection schemes but they will allow people to survive with little hope for the future. Artificial Intelligence will be used across the board in health care, education and across public services. Robe al cops will patrol streets and wars will be fought by killer robots reducing the cost of war for wealthy and powerful but not for the poor. We already know how data driven systems from financial to predictive policing applications can discriminate against individuals. AI systems will become the gatekeepers deciding who can access health care and not. Those are power and access to the fruits of the data economy handful

of countries and companies will be those that continue to gain while the vast majority of people are left behind. This brave new world may entrench global inequality on a scale never imagined. And this will come with massive upheaval and disruption. There is a possibility of another world and this conference and initiatives which follow from here offer as an opportunity to take a critical step in identifying principles for the ethical development and use of Artificial Intelligence.

Companies could take notice and governments could respond. In the future we could have Artificial Intelligence systems that detect and correct bias and data, rather than doubling down on the human bias. We could have automation that takes people out of dangerous and degrading jobs but also educational and economic policies that create opportunities for dignified and fulfilling jobs and Governments could band fully automated weapons system is so that killer robots never come in to existence. This is a future where enormous power and potential of Artificial Intelligence is harnessed for the good of humanity, promoting equality, freedom and justice. It is a future where open source AI allows innovators across the world to harness the power of technology where explainable AI is developed and used allowing for AI decisions to be interrogated and challenged and with clear legal accountability systems to ensure that the rights and responsibilities of users and developers are clear. In short it is a future. AI is a technology where Human Rights principles is the same as a core design and use principle.

But this can only happen if we start out now with an objective to protect human dignity, equality and justice. AI is built by humans and it will be shaped by human values. If we build AI systems that are a mirror to our current societies they will be riddled with historic biases and inequalities of our societies. But we can do better. When the states, they were not reflecting the world that we [live](#) in. But an aspirational world. We must today challenge ourselves to be aspirational again. As we prepare for a future world where AI and technology are integrated in to every aspect of people's lives. Fortunately we already have the Sustainable Development Goals the 2030 goals which almost every country in the world is signed up to the United Nations just two years ago and a core principle of the Sustainable Development Goals is the principle of leaving no one behind.

So this is a principle which would guide us as we think about the discussion here today. And it is heartening to see the interest and dedication to ensure that the development of AI is ethical. We believe that such ethics must be based on Human Rights principles. These are universal principles that have

developed and matured over nearly 70 years and have been applied in national context across the world. Governments have binding Human Rights obligations and corporations have a responsibility to respect Human Rights. We strongly believe that enshrining AI ethics and Human Rights is the best way to make AI a positive force in our collective future. Amnesty International exists to bring about a world that Human Rights and freedom are enjoyed by everyone everywhere. As the world's largest people's movement of Human Rights, today our work is as much about campaigning for the row lease of prisoners of conscious and the protegs of the world's 21 million refugees as it is calling for a future promote freedom, dignity and equality and justice. And so I take this opportunity to announce Amnesty International's Artificial Intelligence and Human Rights initiative, I invite you to collaborate over the coming months and years to ensh rooin the protection of Human Rights in the development and use of AI. So going back to Dinka who I referred to at the beginning of my speech, let's set our mind to think about what happens when Dinka grows up. What happens with our children and grandchildren grow up and how will they judge us on what we did and decided today. Thank you.

(Applause.)

>> URS GASSER: Thank you so much Salil for a fantastic talk. You made many important points. And I hope we get a chance to discuss them with you also after this Plenary. One point you made is to clarify that also Human Rights and the principles have a dynamic nature. It is not a static concept. It is an evolutionary concept itself. And so looking at it from the other way around, do you see some sort of an opportunity there that AI may actually help us to reenvision Human Rights whether it is in the implementation which is more obvious by looking at the principles themselves and that against the back drop against of discussions that we had over the past decade in the light of the digital revolution do we need to come up with new types of Human Rights. Do you see similar dynamics applied in the future for AI?

>> SALIL SHETTY: Sure. There is nothing static about human heights. It is constantly being interpreted. The UN system and other bodies are constantly adapting and interpreting what it means in a very practical sense. So absolutely. But the basic principles, you know, of nondiscrimination of equality of justice of fair play, of humanity, those I think are core values which you in a sense those are the guiding values and those are the ones which you really should think twice before you compromise on it. For example, we take what's happening in Europe around the elections more recently you have extreme sort of populous parties pushing for anti-refugee rhetoric and many

of the mainstream parties then try to start sounding like them in order to norm, to -- that becomes more of the mainstream and I think absolutely you have to adapt to the reality in your context but you don't compromise on the basic principles.

>> URS GASSER: The theme of inclusion is also actually a perfect segway in to the next talk by professor Vicki Hanson who serves as the President of the Association for Computing Machinery or ACM in short. She has many positions. I just highlight two. She is a distinguished professor and codirector of the center for accessibility and inclusion research at the Rochester Institute of Technology and holds a Chair of inclusive technology at Donda in Scotland and led the Foundation at the IBM Watson research center. Vicki I hope you can expand on this key concern, how do we include or create an AI ecosystem that is inclusive. And doesn't leave out the last person. Yeah. Thank you.

(Applause.)

>> VICKI HANSON: Thank you. That was a hard Act to follow. There is no doubt about it. So yeah, you already gave pay way my opening line here. I was going to say that I was asked to say a few words about a C, M which is the Association for Computing Machinery I had a **sdu**gs discussion at lunch. This is not the academy of country music. Those of you looking for that I do apologize in advance. So ACM is the oldest and largest society for computing researchers and practitioners. And even though it is headquartered in New York you can see this is a very global organization. This is a map of where our members are from around the world. And ACM is a volunteer driven society. So the activities that we do are all based on the interests of the researchers practitioners and educators who are or members. Main strength is our technical leadership, and just going to show you a little bit oops, about this here. The clicker is a little -- got a bit of a delay built in here. So we sponsor about 200 conferences a year around the globe and are in collaboration with many more. The proceedings from those go in to our digital library and I wanted to say that ACM has been working to take a lead on a new conference on AI and ethics which has become important topic today. Among those special interest group we were told that ethics is the most critical challenge they are facing these days. That was the motivation **hin** behind getting together a new conference. And as another example of our leadership one of the things that we have is a major awards program and our largest award is the AM touring award. This year it is being given to sir Tim burners Lee for his work in **invent**ing the World Wide Web. This is a very exciting year for us. This is the 50th anniversary of us having give **chb** the Turing award. And as a big anniversary we will be

celebrating the award next month in a day and a half event that has technical discussions from Turing laureates from our own prize winners and from dig dig any tears throughout the field and they are going to be talking about topics relevant to AI. There will be a panel on advances in deep neural networks and one on challenges and ethics in computing.

And I mention this because that event will both be live streamed on June 23rd and 24th, but it is also going to be videotaped and made later. So maybe some of you could go to the ACM website and take advantage of these discussions that we are going to have by a lot of luminaries in the field. I want to thank you for inviting me to this global Summit. Thank you Stephen Ibaraki and the other organizers for having me here. This is clearly a critical time for AI. And I will say that Artificial Intelligence is not my area of expertise. So I was excited to be invited here to learn and I have already learned a tremendous amount. Looking forward to the rest of the event.

The reason that I have found this whole topic of AI for good so inspirational is, for example, this sentence that I found in the report on the 100 year study of AI. Saying the measure of success for AI applications is the value they create for human lives. And this is something that I believe deeply in. My own background is in human computer interaction. Not AI. So I'm particularly interested in how humans and machines interact with each other now and in the future. I am particularly interested in how machines will augment the abilities of humans. My only background is particularly in accessibility research. So I'm interested in people with disabilities. And as you can see in this slide disabilities come in various forms. So this shows the fact that sometimes a disability is a visual impairment. It could be a mobility impairment, inability to move a hand or someone who is unable to walk. It could be a hearing loss. Or it could be a cognitive impairment. Some problem of thinking or processing information. And very typically these days what happens with aging there is combination of these impairments that people may have. So someone may have a visual problem. They may have problems with their hands and all of this makes it very difficult to interact with the world as soon as you get a complex of problems. At a very high level these disabilities create problems with communication, with mobility, with life skills, and with technology. I'm just going to say a couple of words about each of these. So in terms of communication as we all know the ability to communicate is really a major factor of being human. We need to talk to each other in terms of our professional lives it is important. It is important for entertainment. People who have problems with a hearing loss or afascia, cerebral palsy, they affect a person's ability to

communicate and interact with others. Mobility, we all want to move around the world and we want to move independently. So visual impairments in fact, make it very difficult for someone to navigate in the world independently, particularly in a new environment. People who have a cognitive disability have trouble navigating the world because of getting lost. You think of wheelchairs, they are in a place that's not wheelchair friendly it is difficult. Sometimes more surprising is people who are deaf or hard-of-hearing also have trouble going to new places because of the fact that it is difficult to hear environmental sounds, emergency sounds like an ambulance, for example, or it is difficult to be in a hotel because the hotel may not have a fire alarm system that would alert them if there were a fire in the hotel.

And life skills this is just the fact that there are certain tasks that we have to be able to do for ourselves day-to-day to stay out of the care facility. And one of the things that people are starting to talk about these days is the ability to have some kind of digital literacy and interact online with all the Government services that are coming out and that may be a daily activity. The fear that people can't do the basic skills of daily living, people may have to go in to care facilities. That's the normal accessibility think that people talk more often. I am not going to go in to it for a lot of time. That's the kind of thing people with motor disability, for example, use a mouse or keyboard. But one of the things I want to say is I have been using the words impairment and disability. The World Health Organization defines impairment as a health or body problem. So, for example, someone might have a hearing loss. However disability is a complex set of problems and it is actually created by a society. So someone has a disability when they have a hearing loss and somehow can't interact in their environment. They can't hear the danger sounds, they can't talk to someone. They can't hear something on the computer because it is all in sounds. People with a hearing loss are disabled in environments in which information is presented only as sound. And so if we are developing applications with AI or any other way, that are using sound only, then we are disabling people who have a hearing loss.

Okay. Why is this important? I have been talking about disability. It doesn't really seem perhaps that maybe there are that many people with disabilities in the world, right? You probably don't see that many in your day-to-day life. Actually there is one billion people in the world who have a disability of some sort. Why don't you see them? In part it has to do with aging because as we grow older more disabilities do happen but it also has a lot to do with the fact that a lot of the people

in the world who have a disability seem more less hidden. They don't go out in society. Often they are at home. One of the things that I was going to say is that this issue of disability and inclusion relates to the UN Sustainable Development Goal in reducing inequalities, No. 10 over there on the chart. It is a little bit different than the kind of reducing inequalities we heard this morning but there are a huge inequalities for people with disabilities right now. So, for example, people with disabilities experience poorer health outcomes in the general population. Lower educational achievements, significantly lower employment. The employment rates are only about half for the -- for people with disabilities as opposed to the general population. And much higher poverty levels. So there is a huge disparity for people with and without disabilities.

So my goal here is just to do some awareness raising of accessibility among those of you here who are working on AI. So that this kind of disparity doesn't get exacerbated in applications that are built using AI.

I just want to give you two examples about how technology isn't always a good thing for people with disabilities. And then talk a bit more positively about things that can be done. So just to give a first example, I current live in Rochester, New York which is the home of Kodak photography. I used to think that Kodak had to with film. It is mainly the paper that photographs were made on paper. People sent their photographs in and they developed them. Rochester was full of buildings that had dark rooms in them. Who worked in the dark rooms? It was large ly populated by people who were blind. When digital photography came out which was invented by Kodak and they couldn't figure out what do with it. When digital photography came out all those dark rooms were put ou of business and all those workers who were blind and visually impaired lost their jobs. And another story that's sort of similar, probably all know that Alexander Graham Bell invented the telephone, right? And he did it because his wife was deaf. And he actually wanted a way for her to be able to feel sound so that she could communicate better with other people. It had an unintended consequence. The telephone became the greatest disabling technology for deaf people they couldn't have a high paying job because they can't use a telephone and engage in calls. There is a lot of work that goes on in developing technology that doesn't think of the consequences of disabled people along the way.

Okay. So with this background in mind there are some examples where AI has been helpful. Here is current example. This was announced a couple of weeks ago. Facebook, I realize the print is so small. The one on the left we finally made it. If you are a blind person you have no idea what that is about. On the right

it says a Saturday night splurj. Someone who can see can tell what the -- what it means to finally make it. And on the right you can see what the Saturday night splurj is. Someone on Facebook got the idea to use image recognition and now what's happened with Facebook here is if you are a blind person you can get the translation, you can get the image recognition and find out what these images are about. Another thing that's big in disability has always been the idea of independence and I remember that virtually every blind person that I met in my life said they wanted to drive a car. I always said this wasn't going to be possible. It turns out there is some tests going on and it is a little hard it see the but the gentleman in the middle of screen as a white kain and he is blind and they are starting tests so that people who are blind can do the autonomous driving and this is a dream come true for many, many people. I am not going to talk about robots because I'm running short of time. But this is something that's very popular these days. They are being talked about is the future of AI for elderly people living on their own who experience a lot of social isolation. And in closing just like to say -- just like to say, if I get to the right page, that I want people to just think about the idea of developing for a wide range of abilities when thinking about new AI applications. AI discussions often focus on the technology itself looking at things that are faster or more integrated. But these things also need to be accessible and useable to a wide range of people. We have all heard stories about -- this has been talked about before about the bias and datasets. And one of the things that I really hope gets included in some of these datasets in the training is that more people with disabilities have input so that the training will include a wider range of ability.

Last sentence, so other people here I know are going to talk about these data driven and algorithmic biases a little bit more. But consider data from people with diversity in AI development so as not to disadvantage the billion people worldwide whose needs differ from those of healthy young adults. Thanks.

(Applause.)

>> URS GASSER: Thank you Vicki. Very quick question only since we are running a little bit short in time, so given the nuance picture you painted and also the helpful reminder that there are massive in equalities and access barriers. This morning where we were energized with this vision of we will make the world a better place and everything will be great and Democratized. Do we need to engage in expectation management? Do we really have to soften our message or tailor our message?

>> VICKI HANSON: Okay. Tailor our message, yes, I think it

can be a very positive message. This particular population gets left out a lot in thinking. It has been left out a lot in past development and I don't want that to happen again as we start all new technologies.

>> URS GASSER: Thank you. Our last speaker who promised he will keep us right on time, which is amazing, thank you. It is now. It is professor Gary Marcus who is a Professor of Psychology and neuroscience at NYU and published on many different topics ranging from human behavior, neuroscience and AI and he was the previously the CEO and founder of machine learning startup geometric intelligence and he is a frequent contributor to New Yorker and New York Times and author of several best selling books. And it would be delightful to have your final remarks of this panel at least.

(Applause.)

>> GARY MARCUS: I have no slides here. I see myself but that's not what I want to show you. I don't know what's going on with the clicker here. I see. While we wait I just want to say it is an honor to be sharing the stage with some of people who are speaking so incisively about human dignity. And here we go. So I'm going to speak today from the perspective of someone who has worked in the academy as a cognitive scientist, professor for a couple of decades and also in industry as a founder of company that was just purchased by Uber. Wendell tells me I was the first person to write about the driverless car and you can take seriously what I have to say. I don't know. My aim today is to provoke. I am going to start with some contrarian statements and AI and I am going to conclude with suggestions of what might be necessary to move forward if we are going to achieve AI. Where is the field right now? We are not nearly as close to strong Artificial Intelligence as many believe. I think there has been a trend throughout the day where we started with a lot of enthusiasm and there has been a little bit of skepticism filtering through and I will keep the trend in that direction. So here is some optimism from Andrew Yang who was the Harvard business review. If a task takes -- Andrew is half correct, there are some things that people can do in less than a second that AI is good at. So, for example, you can probably distinguish between tiger woods and a golf ball and so can your favorite deep learning net. AI is pretty good at recognizing objects. Typically using convolutional networks that were developed in part by Yoshua and so forth. The basic idea is you have Big Data in and you have statistical approximations out and it works pretty well. I have used one of these apps that does deep learning and this was in my hotel room in Hong Kong over the day. Water bottle and water bottle and third one on the right it came up with pen. This app that does recognition. If

you probe more deeply there is something not quite right. Here u can describe this picture and you might come up with something like a group of young people who are playing a game of Frisbee and you might look at the next one is person riding a motorcycle on a dirt road but if you saw this one you would not say that it was a refrigerator filled with lots of food and drinks. If you were a patient in an Oliver Sachs book but you would not if you were a human being. Yoj la cun and I have these debates. My line is the hallucinations and they are part of where we are right now. Deep learning is got at certain as peks of perception but perception is more than categorization and cognition is more than just perception. So achieve its destiny, AI is going to need to go further than we have already. Whoops. So a good deep learning system might be able to identify the dog at the bottom or might have trouble because we don't see dogs are ears in this orientation and if model doesn't -- real point you can tell that it is dog holding up a bar bell and if you are a human being you have thoit the how did the dog get so ripped. If you were my four-year-old and say who is the elephant Stanning a tight rope. The way I like to think about is it cognitive scientist is there are many things that go in to intelligence, per semgs, compon sense, analogy, language, reasoning and so forth. And what we have made real progress on we meaning Yoshua and not me, what we have made real progress on is perception but the rest we haven't made that much progress in the field. In erm its of understanding what problems can we actually solve in the human fair ta yean domain. In a typical person can do a mental task we probably automate it using AI. What you are doing in your humanitarian organization if you can get it done fast a machine will do it for you. So we have labeled examples. So if you don't have the labeled examples then the techniques are not going to work. You have five or ten. The tek neks are not going to work. If you can do less than one seblg of thought and we can gather an enormous amount of directly relevant supervise data we have a fighting chafrngs. So go is exactly opposite. If you your probably lem is like go we can solve it. You can gather as much datas you a want ab keep iterating over and over again. But if your probe lem is not like go, like it is politics in the real world we may not have the tools for you yet. The stuff I have here in green we do pretty well. Speech recognition especially if it is in quiet rooms with native speakers maybe not so well. We can image recognition is long as the world is pretty bounded. We can do natural language understanding in narrowly bounded domains. Siri can tell you what time a movie is playing and but not answer general questions. Hoich do transistors cost and this is eye lieueding the exponential when the Mac plus was released. Wikipedia can tell you when the Mac

plus was released but no search engine that will come with those two pieces of information together which we would hope for in a general conversational AI system. And AI is great at advertising and targeting but I think that Yoshua and I would like to see AI do more than that. Automated medical diagnosis we are making progress there on the visual side but less so when we have unstructured text from doctor's notes and so forth. We see a little progress. I would love to see how `wem` the Facebook stuff does. My guess it is not that hard to fool. Domestic robots for elder care but the eye is nowhere near strong enough yet. We we like to see driverless cars but safe reliable driverless cars are farther away. Here are some impediments I think to reaching strong AI. First is engineering machine learning is hard and difficult to device. You made modules and bigger modules. And you put them together to make larger mod `douls` but we can't really do the same thing in machine learning and talk at M tech a couple of years ago, Peter Novec will be here by Skype, the way that things work now your data works and you test the model on Tuesday and it seems to work on Thursday and then suddenly it is Christmas and all the `asimions braeshg` and the model doesn't work anymore. We have no procedures for reliably building complex cognitive systems yet. This is by John skully' `es` nephew. The idea you can build the systems and they work in certain circumstances but you don't have those `gearP` tees when you change the rest of the system downstream it is going to work out. XC, this is your machine learning system says one person. You pour the data in and collect the answers on other side. What if the answers are wrong you are just supposed to stir the pile until the answers look right. And here from the ACM is I recently had a piece that was on cover with Erni Davis. Artist due a robot sitting on tree branch and the point I want to make statistics is not the same as knowledge. You have a lot of robots cutting a lot of tree limbs. Y don't what `sto` single robot cut the wrong side of the tree limb fall down and hurt itself and possibly other people with its chain saw. You can't solve this problem with 50,000 labeled examples. Third problem there is bias in the field to assume that everything is learned and I don't fully know why this bias persist but it is pervasive. Yann Lecun he wants to make pour in a lot of video and everything will emerge from unsupervised learning and I wish him luck. I doubt it will work. I was trained in developmental psychology and I have been improving and parityng on the best argue mentes are arguments that say look the brain of the human being starts with something in it. And innate language device and sets of places and things like that in. Nobody every believes me when -- I started making them about baby Ibecs. Here is a baby Ibecs and it is climbing along the hill. If it falls

once it is out the gene line. What building something in. But it is not learning trial by trial in the way that our contemporary machines. We need more innate Psness if we are going to build intelligent agents. We show you our robots which don't have a lot of innate structure and don't do as well. So we lost the sound but that's okay. There is nice little piece of rag time here but you get the idea even without the sound. So we herd a lot about exponentials this morning. I I want to advise a little caution. So in some fields of AI in narrow AI where you have that tightly constrained problem the rules never change and so forth. There is genuinely been exponential progress and you can plot it by looking at scores on chess. There are not no data on strong AI. And here we have Eli za in 1965 and and there hasn't been progress in general Artificial Intelligence. If we are going to fulfill the destiny of AI we want to get the strong AI. There is lot we can do now. There is a lot of fruit to be gained in the next ten years and also a lot of things where we want genuinely intelligent machines. I give you a proposal with three premise. Almost certainly going to -- you go back to my pie chart many different components to intelligence. Many different components to Artificial Intelligence. We need people working together and psychologists and pem working in classical AI and working in neural networks and security and program verification all working together. It is not going to come out of one lab. The human brain is too complex to be understood which one individual and true AI probably is as well. So we are going to need I think a really strong interdisciplinary effort and second thing, in the ideal world AI would be a public good not something that's owned by one corporation or eight individuals or something like that. But we are headed on path where that is what's going to happen. And there are companies like Google that busy patenting very basic ideas in AI some that I think are inn defensible ideas. If you are a startup company is Google going to sue me for using this idea that's twice as old as Google. Google oent uses their patents defensively. They are never going to sue or many of us know that Google has started some fierce intellectual property disputes recently. The world has changed and it could change more. It could change where these eight people who own AI or whatever it might hypothetically be aggressively pursue intellectual property claims and no existing approach to AI approach can get us to next general -- corporate AI tends to focus what can we do with the techniques that came out from Yoshua's lab. How can we commercialize the stuff that's mostly recently been discovered and academic labs seem to work too inned pently of each other. A bunch of small academic may not be enough to get where we need. I am going to make my proposal. The proposal is let's look down

the street from where we stand right now to CERN. It is a global collaboration from thousands of researchers to build technology and science that could never be built in individual labs. Maybe we need to have a model like that for global AI. Lots of people doing AI for the common good. Two more sides. Wendell asked us what is it that what's one detailed problem we can work all on together. I want to give a meta-answer to that. I have been thinking about how to do AI myself and I don't which of the many problems that are out there I should address myself to and what I have been wishing for and maybe someone can help with us is something like charity navigator and not for the individual how should I spend my thousand dollars or five thousands and who can move the levers. People like me who do AI research don't know about. We need some help from other people in the room. Where is there data available. Maybe there is some way to sort of metaize to coin a terrible word Wendell's question. I will send with this wonderful African Proverb. If you want to go fast go alone. If you what tonight go further go together. Thank you very much.

(Applause.)

>> URS GASSER: So thank you so much for a fan fantastic closing talk. We have reached the time for a coffee break. We would like to ask the panelist for a Tweet link statement where you are optimistic or pessimistic about where we are heading to in our field. Who would like to start? Salil.

>> SALIL SHETTY: I am in the business of optimism. Amnesty Internet starts getting pessimistic the world will not be in a better place. It is AI for good but good for everyone and not for some.

>> FRANCESCA ROSSI: I am optimistic because I think that AI will be shaped by humans and I am general optimistic about human causes are.

>> VICKI HANSON: I would say I'm optimistic now. So much excitement and different things canning done and all kinds of create tiff ideas are going to come out of this new work.

>> GARY MARCUS: I am optimistic about the opportunities and inertia won't get us there.

>> URS GASSER: Generally optimistic note with a caveat. Let's thank our speakers. Let's have coffee.

(Break).

>>.

>> ANJA KASPERSEN: Could I ask everyone to please sit to we can get started with this exciting session? Thank you. So my name is Anja Kaspersen. I work for the international committee of the Red Cross where I am looking at various ways that AI can be used in humanitarian activities and deliver humanitarian action and how it is being used in warfare. With me I have my

co-Moderator Nicholas Davis and hopefully we will create a flow in bringing in a lot of interesting and powerful voices on the issue of AI and how we move forward on some of these issues and opportunities we have been hearing about today.

>> NICHOLAS DAVIS: Thank you. And thank you everyone for being here. I'm Nicholas Davis I lead the World Economic Forum society and innovation work. Focusing on innovation and they come together in the discussions we are having today around Artificial Intelligence for good. I am delighted to be here with Anja and fantastic [roun](#) of speakers on stage and then a panel discussion to come to discuss the how. So we met this morning in the opening Plenary and we talked about the current state, the current state of play for Artificial Intelligence. And in the Plenary just come, just been, we spoke about the what to expect from the future, how there are certain aspirations and concerns about what may unfold but the most important part of the next two days and this evening is to focus down on what we need to do. At an Artificial Intelligence algorithms are being developed and released and embedded in to products. How can we come together as a global stakeholders to think not just about people who are lucky enough to come to a global meeting like this in Geneva but the people that Salil Shetty talked about in the last Plenary and people who have less voice and with those people that Anja is often concerned and working with at ICRC.

>> ANJA KASPERSEN: Indeed. When talking about solutions as we are trying to do during this conference it is important to develop a shared narrative where we address whatnot just what is at stake but what are we trying to address. What are we trying to address for whom by whom. The impact of AI alone are in convergence with other technologies which is one of the [esh](#) issues that I work with. If you on-the streets fleeing a war in Iraq or standing in downtown Geneva trying to hail a cab. So addressing the as much as there are opportunities there is also risks and hopefully we will get to some of those risks and dilemmas and controversies that we have to address to ensure that using AI for good and we preserve those that are the most vulnerable when they need our protection the most.

>> NICHOLAS DAVIS: In this session you are really getting two Plenaries for one. This first section we have five fantastic speakers. We are going to start with Robert Kirkpatrick and we are going to move with em-Katsumi Emura Manuela Veloso, Cindy Smith and we Stuart Russell. Importantly to frame the discussions over the in ex-two days we move in to breakout discussions and more focused groups for thinking about what it really means for Artificial Intelligence to Act for good. So with that thank you Anja. You will see Anja who will come and bring our panelists on stage for the second half of this

discussion. I want to make one moment and say we are in Switzerland and I have lived here two years. Anja will bring us to close at 1815 this after noon. We are going to focus on crisp and concise perspectives from our speakers and thought provoking discussions from the panel. Please don't think of me as Nicholas Davis as the Moderator who is human and understands things like social norms. Think of me as someone who is an Artificial Intelligence agent with an iPhone timer in their hand whose purpose is to make sure we get to drinks and enjoy each other's company and discussions after two hours of fantastic conversation. And with that I want to invite Robert up to the stage. Robert is not just the person who runs UN Global Pulse, the Big Data focused aspect of the United Nations. But also is an entrepreneur, a social entrepreneur in many senses and someone here who can let us know what it is that we need to do particularly in the the use of data with regard to Artificial Intelligence. So Robert please welcome to the stage.

(Applause.)

>> ROBERT KIRKPATRICK: Many thanks and good afternoon to all of you. I was asked to speak, of course, today on aspects of a roadmap for Artificial Intelligence for social good. Given that although I'm an optimist with AI, sort of is this an ostrich or a pekine *se* I am going to be focusing on narrow AIs and thinking about that roadmap as a start. So we live in a *world* in which everything increasingly is smart and everything we do emits data as a biproduct of our actions. At Global Pulse we Big Data and all this information out there is as a new natural resource. It is abundant and renewable through the power of Artificial Intelligence and high performance computing and can be *refoined* in to new forms of information. Tremendous value for answering business questions and for fighting hunger, poverty and disease. Our labs are using Big Data and deep learning to map discrimination against refugees in Europe and recognize rescue events from shipping data in the Mediterranean and detect fires in Indonesian rainforest and predict food prices from twitter, fight bottlenecks in sanitation services in Satari refugee camp and generate socioeconomic indicators from satellite imaginary and flows of postal traffic. Lots of data and algorithms that can be used for good. We have heard a lot about speech technology today. Talk radio programs are the oldest and most original social media platform out there and still the primary one across the world. We are using to generate realtime insights from public reremarks on FM talk radio shows. No one had grasped the potential for spotting things like cases of Ebola or small scale natural disasters that would otherwise go undetected. These are small examples compare to the advanced institutions in the wealthiest part of the world are taking. Our

new natural resource has fallen in to the hands of an extractive industry, one that is opaque and unregulated. Its benefits are not reaching those who need them most and the expertise needed to refine these material are and you affordable to all but the wealthiest. And finally this resource is a bit like nuclear energy. It is ha a habit of leaking sensitive information in ways that can contaminate and cause harms for years. To how this information can be used in cyber warfare tools and autonomous weapons. In 1950s when we had seen the horrors of atomic bombs but had scientists tell us that we can use this to safely power science for years safely. How do we contain it safely and harness its power for good and how do we govern it: There is analogy that we can build on.

But that isn't only challenge. It is not just leaking. We know how hard it is to [anomize](#) this data and also know that bias can be introduced in to it. Either by the data used to train it or by the purposes of the developer who originally designed an algorithm. There are huge challenges in transparency and we heard some of that today. Maybe the best efforts we will have will still have an algorithm telling us why it said a decision. At the end of the there is a profound lack of trust and we are facing issues with it where we don't know what the next step is. Rupert Stadler this morning spoke about self-driving cars making a life and death decision. What if some portion of the society say we think that algorithm should have access to everyone's Facebook pages, how people feel about this and what they buy with their credit card so they can evaluate the value of their life. How do you make a decision and some portion of people will say we don't want machines making life and death decisions about the value of human life and yet we have having to program that right now and others will say if we don't it is not fair. What about privacy. What makes this data and these technologies so powerful also makes them dangerous. It is not easy to separate. Nevertheless though the opportunity here is unique. Even the relatively narrow AI applications we have heard about today have the opportunity to be used to change the world. It is probably clearer to all of you that the ambitious nature of the Sustainable Development Goals means we are not going to be able to achieve them unless we transform how we make decisions and have massive increases in productivity. Looking to define a roadmap we have to first agree on where we want to go and while we are at it we should make defining explicit nongoads part of that process. The next step is ground our roadmap and principles contained. So basically access to health care education, nutritious food, clean water these are Human Rights. Our approach on the one hand must accelerate adoption. AI must as a tool of empowerment of everyone on earth but on the other hand,

it must also adhere to the principle of first do no harm. We aren't doing well at this today. I would argue we need to make sure that AI gets diverted from the path that it is currently already well on. Which is increasing inequality all the around the world. At the end of the day we have navigating a course of risks of using these new tools and risk of not using them. In our own work we are encountering cases where a particular opportunity to do good would be *ral* valuable but we can't do it because the risk of misuse is too great. That is only going to increase as we look at the applications of AI. So a roadmap I think divide the roadmap in to two categories one is how do we seize the opportunities and the other how do we mitigate risks. The one could start with a fund. Imagine a global fund for AI. Establish potentially as an international institution within the UN. This fund could invest in cutting edge research and promoting AI based tools directly addressing the most pressing challenges of the SDG. And scaling proven solutions. Ideally through open source tools where possible. We heard earlier I don't know where to invest. I don't know what the best ideas are. Maybe such an organization with a partnership for AI could help define a common agenda for where the most *ksh* -- the highest priority investment should go. Government should invest in building the capacity of their public sector institutions and civil society organizations and promote computer science to increase the pool of talent out there. We are going to need them in an operational capacity all day long as we start to build these systems. I would also say, you know, on the problem of we need to make sure we are investing and not increasing inequality, of course, and, you know, if you look at the opportunity if you look at the challenge of preventing automation from taking jobs, I'm skeptical that simply giving people universal base ic income solves the problem. Finally we know we are going to need a multi-stakeholder process to get to principles of AI. We could ultimately feedback on how well that's being implemented, we could see international soft law be put in place and one day perhaps a Convention. I look at the existing rights instruments that are out there today and I don't believe that any of them will adequately address all of the risks and opportunities around AI. In particular when we move to a generalized AI. Last point simply is that I think that what needs to happen next is a series of demonstration projects that can show the value of Artificial Intelligence at the -- and its cutting edge state in ways that the public can understand that also demonstrate the very real and frightening risks and ways this can be mitigated. If we don't address the trust problem we won't be able to move forward. Thank you.

>> NICHOLAS DAVIS: Thank you so much. Stay up here for a

second. As I am sure with all the panelists and speakers there are a hundred questions I want to ask as a follow-up. One particular one, do you have an idea of a demonstration project that this group of multi kind of faceted stakeholders from around the world should start to get behind or design over the coming days that can do what you mentioned? Illustrate the potential but also the potential for misuse?

>> ROBERT KIRKPATRICK: I think one project that could be done, for example, would be learning -- I'll give you an example because we are having some discussions about this sort of thing and it was raised earlier by Dr. Margaret Chan this morning. What if you could bring together all the data on what people are buying and selling, the movement through cities, particulate information in the air and distances to parks to gyms, to McDonald's availability of biking lanes in cities and map down to the neighborhood level the risk of noncommunicable diseases. To say if you live in this neighborhood these chance of ten kinds cancers, because of the built environment and owe die bet **tets** you could begin to understand. If you live ten minutes away your children have a much lower chance of dying of Leukemia. It could change real estate prices, insurance rates. It will put huge pressure on mayors around the world to alter the built environment. It would empower citizens in a way they never had this before to understand what it means to live where they do. The only way to do that is bringing this data together in a privacy protected way and running a lot of deep learning on it to be able to understand the risks. And then we need to deal with the consequences because the truth sometimes is a problem.

>> NICHOLAS DAVIS: Thank you very much Robert. Please thank Robert Kirkpatrick from UN global pulse. Please.

Robert you can go and sit back down. Otherwise she will be left alone at the panel and it will can like an episode of Survivor of some sort. Professor Manuela Veloso. Thank you for taking the time to come here and join. Professor Manuela Veloso is the Herbert A Simon at the school of computer science and leads the machine learning department. If you ever Googled Manuela you will see the exciting work that got me interested in augmentation of humans of Artificial Intelligence is the work of the cobots and robots. If you are not going to talk about I will leave that for the audience to talk about. Thank you very much.

>> MANUELA VELOSO: Thank you. I'm going to tell you a little bit about another form of AI or somehow the form of having robots that move in our environments. If I could have my projection. First time I'm giving a talk in which someone is projecting what I want to say. But that's okay. So what I'm going to tell you is this thing about we kind of understand AI in terms of cell phones and computers and a little bit drones

and Internet of Things but I really want to talk a little bit about what is this thing about having robots that move around. Robots that move combine the ability to perceive the world. They have sensors and computers that enable us to think and they have action in which they move. This is robot, cobot, collaborative robot that moves at Carnegie Mellon of performing actions and has moved at CMU. In this talk I am going to very briefly in the eight minutes that I have to tell you about three things, three things. One is in fact, that these machines that move around are need to be able to localize, need to know where they are and they need to actually plan for routes and execute them.

And then I'll go on in to two second topics. So in this first topic, if the slides move, let me just bring you to the core of what is robots or these AI machines are all about. So they have sensors and these sensors are not our eyes, not our touch, are not our ears. I mean it is a different world. The robotics world they have sensors and in this particular case I am giving you an example of an an Lidar which is a scan of rays that tell you how far the obstacles are. And, for example, they have a 3D camera a that gives also what is like the scene in 3D with very limited kind of distance, six meters for the regular connect or maybe can get ten meters and you get these obstacles which are the walls. With this information they could have other sensors but these are common sensors in robotics. With this information they basically have to be able to transform the image which is on the left side which is a depth image in to eventually mapping whatever they are seeing the distance at which they are seeing walls to a map that they are given or that they learned. So somehow robotics and a mobile robot has to map whatever they see with this very limited sensors to whatever they know which is the map of the environment. So the way they function it is actually through these -- oops. The video that I was going to show, so they are supposed to map and they actually move at Carnegie Mellon as you saw before, moving four kilometers. They have move for more than a thousand kilometers. If you come to Carnegie Mellon tomorrow you will be escorted about I a cobot. Cobot is escorting some visitors which I scheduled yesterday and it is escorting visitors today. You don't come do to my offices by knowing my office but a cobot and you follow the robot that takes you to the office. And which it does. So that's very interesting kind of phenomena. These robots don't have ARMs and it is kind of like understanding that is like some limitations that are embedded. They don't go upstairs. It might be hard for them to open all the doors of the world. So it becomes this kind of like realization that our robots have limitations. So back about five years ago these type of like realization led me and my students to introduce this concept of symbiotic autonomy.

These robots ask for help. So these robots are left in to the environment. But when they find something they cannot do they don't understand, they are lost and certainty is too high they just stop and ask for human to generously help them this pressing elevator button and telling them where they are and put things in their basket. They navigate autonomously in this kind of symbiotic with the other robots because there are four of them and with the Web. So if the next video works, I'll show you an example of these in which if you put the sound -- well, you can tell the robot please bring a coffee to the left and the robot says I don't know this object. What is coffee. And I'm going to check what's probability of coffee being in these locations.

>> Also for object kof which location do you think is best.

>> MANUELA VELOSO: Human says go to kitchen to get coffee. This robot goes to kitchen and he is going to the kitchen could get this coffee. Now I can ask you how is it going to get coffee from the kitchen given that it doesn't have any ARMs. It can't do anything like getting coffee. The way it does it gets to the kitchen and talks to the air and has no clue in there is people there and it just says hello. Could you please put the object coffee in my basket. And then there is someone in the kitchen that very generously actually does that and puts coffee in the basket and then says okay, now I can go. So there is this kind of like now we can have one day the robot have ARMs and it might not ask any more to put things in the basket. But up to now all these things it can't do yet and many of the things it has been improving on it asks for help and it gets to the lab and says I brought coffee to the kitchen. Tell me what I do and there you go and you are done. There is this new paradigm in which the robot asks for help. So the important thing here to understand in some sense which is novel with respect to other things in which robots had like being supervised is that this is proactive ask for help. They find out when to ask for help which is specific generalized situations and then eventually they substantiate them as needed but it is not that anybody supervising this robot and then finally -- and they actually learn from this interaction and I want to tell you something that's important. So those robots started knowing nothing about which objects were where or where Manuela's office was and everything was learned. Hundreds of facts were learned through this interaction with humans, human machine the kitchen and then it saves mapping and this robot has acquired all this knowledge about the environment. We can instruct it in policies. One final thought when this robot appears from front of someone's office and this is the last difficult problem about robotics and AI, you wonder you did not see anything. It just stopped in

front of your office and there is this nightmare about finding out what happened. So when cobot stops in front of my office I want to know what happened while you were on its own. Did you escort successfully Stuart back to the elevator. Did you -- what are you going to do next. How many times you go to the kitchen today. So the autonomy and this AI has to -- we have to realize that there they are going to do things on their own that we actually don't know about. So we have introduced this concept of verbalization in which the robot can generate automatically the at the description of what its did, I went straight for 15 minutes and I turned left and then there is verbal ization to generate explanations of different levels of ab straktion and the robot can give different levels of detail in its explanations. So in conclusion the three points I wanted to make, if this goes, so the in conclusion, so there is this great challenge or this great opportunity for researching developing and really producing things on human and AI interaction. And we have been doing this on humans and web and robots this coordination and this task distribution but, of course, hopefully for the joint benefit, co-existence and transparency which is something that we are all researching on to have these machines that learn, and they have to do things auk automatically explain themselves when humans express explanations. Thank you so much.

>> NICHOLAS DAVIS: Thank you for summarizing what has been more than 7 years of research in eight minutes. But here you have an audience of stakeholders from civil society, from the tech world, from the UN, of course. From other industries. And I guess the question that's burning in my mind is what new areas or extended areas of research should we be thinking about or investing to extend the lessons that you learn about collaborative robots in to helping a broader set of stakeholders and not me or us when we come to visit you at Carnegie Mellon. I.

>> MANUELA VELOSO: I do make the case let AI be mobile. I entered this building there what is not a single robot moving around itself. Go to hotel and there is no robots moving around anywhere and we are surrounded by AI technology that does not move. So my first point is like if you can in your companies believe that in fact, this mobility as possible and then the second, I mean the second case is in fact, let's acknowledge that AI systems are not perfect. And it was going to take a long time for them to understand all types of scenes and deep learning. There will always be something that they don't get it. There will always be something they cannot do and what about thinking about these machines as being able to ask for help and say I need more data. I cannot classify these any more. I need to ask another -- I need to ask a human. So these -- so the

first thing is mobility and the second thing is acknowledgement that we can have AI for good. AI deployed for great results even if we have not solved the `whoelt` thing yet. If we enable our systems to ask for help and declare they need more knowledge and the third is not make these black boxes and always try to have the systems be as much explainable as possible. And this is a research question. And so that's basically the three things that I think we can all work together on is these aspects. Okay?

>> NICHOLAS DAVIS: Thank you so much. Put your hands together for professor Manuela Veloso.

(Applause.)

>> NICHOLAS DAVIS: And now it is my pleasure to introduce Mr. Katsumi Emura who is the executive vice-president and chief technology officer for -- we are shifting here to thinking about the roadmap on a national level and fascinated to hear your views of what we can learn from Japan's experience here and where we can take it in to other countries. Please.

>> KATSUMI EMURA: Thank you. Good afternoon Ladies and Gentlemen. As Nick introduced me I'm from industry. And actually EC corporation but today I'm presenting Strategic Council for AI technology in Japan. That's today. I introduce Artificial Intelligence industrialization roadmap developed in Japan for realizing society 5.0. I will explain society 5.0 breakdown. Can I get my slides? This is not mine.

>> NICHOLAS DAVIS: That's Stuart Russell's slides.

>> KATSUMI EMURA: So when we discussing or developing roadmap there are two ways. One is on kind of forward casting. So in a technological people always use the forward casting type of roadmapping. But in industrialization roadmap in Japan we are using broadcasting approach.

So in the broadcasting approach we have to advance in the society in the future. This is one example. This shows a global state in 2050. The population in 2050 is expected to be 9 billion. Maybe 30% increase from today. But because of the urbanization the people living in the stay will be increased in 1.8 times. Resulting in `hoouj` huge increase in resources like energy, water and food. How to cope with demand is our challenge.

And the SDG, the 17 goals, this -- these are also showing our directions. And NEC is committed to contribute to SDGs. And -- but we have seven themes for value creation. And the -- these themes are pretty much aligned to the SDGs. The question is how to -- how do we create social values and roadmaps that align with these targets. And to create social values we have to use AI technology in the sense of visualization, analysis, and prescription.

And the -- by considering these we are developing the industrialization roadmap. And the -- this explains society 5.0. The society which we want to realize in the future. The society 5.0 is a Japanese vision for upcoming society. Following on from hunting and gathering society that is society 1.0, Agrarian society, industrial society and information oriented society. Information 5.0 refers to a social mode that realizes the advanced fusion of cyberspace and physical space. The second provide good and services regardless of location, age sex language or any other indication. And thirdly bring about the human centric society in which people can lead high quality lives full of comfort and vitality. So realize society 5.0, evolution of Internet of Things and Artificial Intelligence are the keys.

Japanese Prime Minister announced April last year saying that Strategic Council for AI technology is established to develop AI roadmap by gathering top experts from industry, academia and government.

The aim to create AI industrialization roadmap is to promote the social issue Resolution and enhance the Japanese industry competitiveness. First we define social issues to be solved. Then we envision the future society. For example, as I mentioned to overcome the resource requirement we must realize, for example, 0 waste society. Under this vision we develop the roadmaps.

There are many potential areas to develop a roadmap. But the considering the next step, means that to realize research project, those kinds of things. We try to pick some of them to realize the roadmap. The criteria were really challenging social issues. Contribute to achieve Japanese SDG target and AI is really contributing. And then we selected three areas. Those are productivity, health, medical care, welfare, and mobility. (GDP)

And the -- this shows the vision for the society under these three areas, productivity, establish an ultimate ecosystem that will enhance productivity in society as a whole, meaning from production to consumption. The health medical care and welfare create society where people at the age of 80 able to work actively. We may redesign social and industrial systems. Mobility, get rid of accidents caused by human error. And this morning we heard that the 90% of accidents caused by human error. So minimize social cost of transportation and the -- when the autonomous driving has become reality we could use moving spaces for workplace or entertainment.

When we develop the roadmap, we have technology trend in mind like semi-conductor architecture for Artificial Intelligence, the kind of data, and the multi core architecture and their

functionality. This shows the whole aspects of roadmapping activities. As I mentioned we see social challenge and technical trend. And the -- with that we like industrial -- industrialization roadmap in a phased manner.

So this shows the phase. The phase 1 is approximately to year of 2020. We already use Artificial Intelligence but in an individual domains. In phase 2 to be around 2025 to 2030 where we apply -- we can use AI's cross domains. So the different domains are connected. And in phase 3 Artificial Intelligence as a service becomes common and with that any industry is connected easily and wholistic solutions are becoming available.

This shows the whole view of the industrialization roadmap. Each squares indicate not the technical goal but the indicator, what we want to realize in the society. So our roadmap is indicating as such.

So this is one example. Maybe you cannot read. And the Japanese materials are already issued end of March and good news English document will be available soon. So and you can read it and check the detail.

So in conclusion we are seeing the society 5.0 and the create the industrialization roadmap. And then next step maybe we do innovation program that is research program and to realize social barriers. I believe sharing the vision and roadmaps are effective way to align activities and good use of Artificial Intelligence technologies.

So thank you for your kind attention.

(Applause.)

>> NICHOLAS DAVIS: Thank you very much. So this is I'm sure fantastically detailed but also also represents a lot of discussion and debate amongst Government, industry and social partners as well. Can you maybe just briefly let us know was there a particular technique or lesson about how to bring together those stakeholders across industry and Government and civil society to come to a common vision for Japan's future with Artificial Intelligence?

>> KATSUMI EMURA: The key is to start discussing the future issues. Not the technology. And the -- for example, like on mobility, to reduce the human error based accident. Then we start to discuss what technology can contribute. Then we start discussing autonomous driving or mobility system. So with that different stakeholders can discuss on the same issues. So that's a way we do it.

>> NICHOLAS DAVIS: Thank you very much Mr. Katsumi Emura. Thank you Katsumi Emura from NEC and the Strategic Council on Artificial Intelligence. You can see there is a trend starting to emerge in the discussions about setting social goals and understanding some of the technical aspects about how our

society and human aspects relate to the technological capabilities and data and other aspects below. I want to hand over to Cindy Smith who is the director of UNICRI. United Nations Interregional Crime and Justice Research Institute and Cindy mentioned and will mention I'm sure the fact that the center that UNICRI has announced a new center focusing on Artificial Intelligence. So it is a delight to have you **hooer** and we look forward to your remarks on what **whe** need to take care of as we think about the impact of AI on multiple parts of future as well as, of course, crime and security. Thank you.

>> CINDY SMITH: Thank you very much. Just in case I don't get all the way through my message in my eight minutes I want to tell you the important part first and then I will tell you how I got there. We were asked to talk about how, what do we have to do and what we believe at UNICRI has to be done is the Cs, my name is Cindy. So you will remember this. The Cs, conversation, collaboration, cooperation, coordination of and as you just heard from the last speaker, private industry academia and Government. Putting those three partners together at the table and making sure that they have a conversation that they collaborate and that they cooperate and that they coordinate.

With that in mind Distinguished Guests and colleagues and friends thank you for inviting me here today. It is an honor to be here. In my short time with you I want to touch upon three key issues. I will talk about first the benefits of AI. Second the risks associated with it and third the necessity of international cooperation to maximize the benefits and minimize the risks. The first of these issues, potential of AI to bring about benefits to global development and societal change has been the topic of discussion so far. We are living in truly an extraordinary moment of history. The world nurtured by science and technology has become increasingly interdependent and interconnected. Rapid advancements in the field of robotics, coupled with the rise of computing power have exponentially increased the breadth of task that can be assigned to **ro** bots and systems based on Artificial Intelligence as well as autonomy with such technologies operating. We are only beginning to understand how these advancements can be really be applied. We at UNICRI believe the benefits of AI will **P** of paramount importance to society in the years to come. Of the United Nations Sustainable Development Goals. For instance, AI may be used to identify and correct for gender bias in the labor market. A wide range of critical fields of relevance to the SDGs such as health, education, energy, economic inclusion, social welfare, the environment as well as crime prevention a topic close to my heart, security, stability and justice.

All stand likely to benefit from the progress made. The

second issue is that advancements in AI will lead to serious challenges and risks. The wide range of economic, demographic, environmental, legal and political implications of these have yet to be truly understood. Looking at some of the challenges and risks we should start to think about what new threats to international peace and security will emerge from the advancements in AI technologies. If you build my folks will use it. And my folks are people who usually end up in corrections. In fact, some of these technologies have already been used for terrorist purposes. We will -- what the -- what will be the availability of new instruments of terror mean more to the international community as it strides against the tides of terrorism. Are we ready for this kind of terrorism? The recent increase in cyber attacks takes a whole new dimension. We must capitalize on AI in this fight against crime and terror and find ways for AI to contribute to enhancing global security.

Moving beyond, a strict traditional security context we see even more challenges and risks. We see, for instance, that AI is beginning to disrupt the global workplace. Change the types of jobs available, and reshape the skills that workers need. This leads us to the question how these changes will be felt in the developing world and what it will mean for labor migration. Whether AI leads to unemployment, increases in inequality or new waves of crime and terror, will depend on a proper understanding of the global implications of AI. Having looked briefly at both the benefits and risks I turn to the third and final point that I would like to emphasize and that's the importance of the Cs that I spoke of. The cooperation, as we strive to strike a balance. We in UNICRI believe that given the technology complexity and the rapid rate of development in this field we can only succeed to minimize the risks and maximize the benefits by bringing together the various communities unifying national and international public and private academic and practitioner communities. In 2015 UNICRI launched its Artificial Intelligence and robotics program. The first at the United Nations General Assembly. Firmly positioning cooperation as being central to our approach. A High-Level Event with preparation of world renowned experts in the field of AI on the 7th of October in 2015 at the 70th session the United Nations General Assembly marked the beginning of this program and some of you were there. Since then UNICRI has sought to advance understanding of the various aspects associated with developments in AI and robotics and contribute to building bridges between stakeholders. We have organized a number of high level gatherings, period of times meetings and training sessions. The most recent of which took place at the University of Cambridge in February of this year. To bring together a group of stakeholders to discuss the **risics**

associated with AI and robotics and how we can leverage these technologies to the benefit of the people, countries and societies of the world.

We believe this work is essential. Societies have to be ready for the technological revolution. In this regard we must strive to avoid stifling innovation and work to contribute to the increase of the human well being. During a High-Level Event at 71st session of the United Nations General Assembly in 2016 I announced that UNICRI was in the process of opening a center on Artificial Intelligence and robotics in the Hague, the Netherlands to serve as an international resource on matters related to AI and robotics. The process of opening the center is nearly completed and for this we are very grateful to all of our supporters and stakeholders. We are particularly grateful to the Government of The Netherlands and the city of the Hague for making this historic initiative a reality and enabling the creation of the global center to discuss advances in this cutting edge field. The aim of the center is to enhance understanding of the risk benefit duality of AI and robotics through improved coordination, knowledge collection and dissemination and awareness. We want to leave no one behind. Our desired outcome is that stakeholders such as policymakers, and Government officials will possess the knowledge and understanding of both risks and the benefits of such technology so that these leaders can contribute to the discussion on these risks and potential solutions in an appropriate and balanced manner. Such solutions may, for instance, take the form of improved policies. The knowledge and understanding conveyed to policymakers should be founded upon real and accurate information derived from the technical experts in the field. A series of activities have already been identified to facilitate the production of the outputs and to contribute to the objective of the center. For example, risk assessments. Training and mentoring programs and so on.

In conclusion with more than 50 years of experience UNICRI has structured its work programs so as to ensure it remains open and proactive to the needs of the international community. Now we are ready to take up the gauntlet with respect to AI and I truly hope that you will feel encouraged to share your expertise with us and exchange ideas with us because only together can we make this rise to meet this challenge. Abraham Lincoln said you cannot escape the responsibility of tomorrow by evading it today. This conference I believe is the best way to start shaping our future and the present. Thank you.

>> NICHOLAS DAVIS: Thank you so much Cindy. I have one I think very important follow-up question which goes to all of your Cs. In the 50 years that UNICRI has been bringing together

Intergovernmental agencies and partners from very different part of the world to tackle incredibly important issues, what have -- what has UNICRI learned and what have you learned as director that can unlock this power of cooperation. What kind of one lesson or tip can you give us about bringing people together to make real progress?

>> CINDY SMITH: This is my favorite question. We think about this at UNICRI in a triangle. And if you think about academics over here and I'm a retired University professor, so pardon me if I say something mean about me. But academics speak to themselves usually and they talk to each other. And they don't relate very well with the other parts of the triangle. And over here we have practitioners and many, many many, many years ago I worked in corrections and so I was a practitioner. And I know that we spoke a different language. So think about academics speaking Spanish and practitioners speaking French. They are the not going to make the change in the Government. It is the people this at the top and the policymakers that make the change and they speak Italian, for instance. They have words in common but they don't always mean the same thing. UNICRI sits in the center of the triangle. And we are the interpreters between the language. So when the corrections people say I don't understand why you want me to do this. And the academics say but I have research that demonstrates you will do better if you do it. The policymakers have to be able to say we are going to make a policy that works but they need both of the information from both of those legs in order to make good well informed policies. We will interpret and you guys sit at the table. Thank you.

>> NICHOLAS DAVIS: Thank you so much Cindy.  
(Applause.)

>> NICHOLAS DAVIS: Now speaking of academics who speak multiple languages we have here professor Stuart Russell. Stuart is someone who you may have seen speaking on these issues before and I am glad this is not my first time of being up on stage with Stuart but you are not just a leading light in AI research on the goals of AI and Artificial Intelligence and you obviously teach a variety of courses on this. But you also work on the applications of AI particularly in medical fields in a sense of computational biology and other areas well. You can bring us home with the mix of technical language and the idea of where we need to take AI from a technical and design sense and what impact that has on us as people who speak multiple different languages around the world.

>> STUART RUSSELL: Thank you very much. It is a great honor to be here and now we can have my slides with -- so Lisa Doll is one the great go players of the world and here he is having what I my Silicon Valley friends call holy cow moment. He

is realizing that AI is progressing faster than we expected. Coming down the pipe we are going to see, for example, technologies that can read. That can read text and understand it extract information in a useful way. Once that happens, very soon after that machines will have read everything that the human race has ever written.

So now you have systems that can look further ahead in to the future, like Alpha go looks further ahead on the go board and they know more about the world than we did and if they can combine those two capabilities then it seems inevitable that AI systems are going to be able to make better decisions than humans across a wide range of activities.

And there is a big upside and people have mentioned the upside. There are many things we can do. First we solve AI and then we use AI to solve everything else. So just to give one example from my group we made some development in AI and we were able to use that to solve the problem of monitoring verification of the nuclear test ban Treaty which is run out of Vienna.

And basically everything that the human race has that we value is the result of our intelligence. So if we can have more of it, then this will be a step change in our civilization. But at the same time there are these risks as other people have mentioned, Cindy, for example. And so we see headlines like this. So all of a sudden we find that AI has been manipulating the information environment that everyone lives in. So it is not -- this isn't our physical security. This is our mental security. And I actually believe that we have a right to truth. There is a human right to live in a roughly true information environment. Because humans are defenseless against information environments that are grossly corrupted. And I think this is a very important thing and it is messing up a lot of our political and social processes right now.

So what do we do about it? Well, Cindy talked about malware, the use of computers by criminals to do crime. That's a very serious problem. When we add AI it is going to get much worse. And the use of malware has been a catastrophe. The world has essentially completely failed to defend its populations against malware. So direct theft from bank accounts is now over 100 billion dollars a year.

And almost nothing is done about it. There is almost no traceability. There is no cross-border enforcement. So we have a really serious situation. We are about to lose the Internet. Large corporations are no longer using digital payments, for example, because it is too insecure.

So I think we need to work on malware and I think by solving malware we will set up a paradigm for how we can start to think about controlling the misuse of AI. But we have to do this soon.

Another area of negative use of AI is in killer robots or lethal autonomous numbers as United Nations calls them and the issue these will become a weapon of mass destruction. Because the thing about autonomy is it decouples the number of weapons that you need to launch from the number of people that you need to launch. The CCW is moving forward with the Treaty. That's very good. I think we need more public awareness. People need to make movies about what this would be like for human populations and our professional societies need to agree that we are not going to do this. Building machines that can decide to kill humans is not okay.

Another risk that people talk about is employment. And this is a very serious question. Economists many Nobel Prize winning economists think this is the biggest threat facing the world economy in the next 20 years.

So what can we do about it? One of the things that you hear is well, it is not going to happen. Robots are -- they are not smart enough and they are not going to replace human jobs. I don't think this is a wise strategy. You could do some careful economic analysis and this has not been happening. The thing we really need to do is have a destination. And we all talk about having transition plans to help our population through this big transition. You can't have a transition plan unless you have destination. What is an economy that you would like to live in where most of what we call work now is done by machines.

And a hint, that economy does not have a billion data scientists working in it. Okay? That's a complete nonstarter. That's a tiny lifeboat for a giant ship and that's not the future.

But this is what I actually want to talk about. So if we have AI systems that are more capable than human beings. One of the risks that people talked about it sets up a terminal conflict between the human race and machines. So people say it is not going to happen. People say don't worry about machines becoming more intelligent than humans. We can always switch them off. That's ridiculous. That's saying we can beat Alpha go. You will find the AI community in denial about this question. The AI community is now saying after 60 years of saying we are going to succeed they are now suddenly saying we are not going to succeed. We have wasted our lives because AI is not going to happen. This is an unwise strategy. We have to anticipate and avoid the fail **kwur** modes. We better be quite sure the purpose put in to the machine is the purpose that we desire. This is the primary failure mode. Intelligent machines will be incredibly good at carrying them out what we ask them to carry out. If we ask them to carry out the wrong thing we will be sorry and this was a point known to king Midas the legendary king it Greece

that asked for everything to be turned in to gold on everything he touched. We have a long history of failing to specify the purpose correctly. So this is a technical problem and this is the main source of risk from AI. So there is a version of AI which I'm calling human **compatdible** AI which is different from the way we have conceived of AI up to now. The first principle is that the only objective of the machine is to maximize the realization of human values but the second principle is important. The robot does not know what those are. It is precisely the single minded pursuit of an objective that causes a problem. When the machine is uncertain of the objective but obliged to help the human achieve it you can switch it off and so on. It will ask questions. It will hesitate before doing something and ask if this is the right thing to do. And then the source of information about what humans want is what humans do. This is a long story about why that's actually right. So I'm not going to go in to that. So my recommendation for what we have to do is actually change the very definition of AI and the way we pursue the research in AI. To make it provably beneficial. So this is an oxi moron. Beneficial is a vague notion. That we have provable rock solid cast iron mathematical guarantee that the systems we build will be beneficial and the recipe is the recipe that I just gave you. We have made some progress on the basic technical principles. And we have actually shown systems that are provably beneficial in the sense I described. As this technical work progresses we need to roll it out through education, through standards, eventually not now but we will need regulation for the design templates of AI systems that are allowed to be deployed and connected to the Internet put in to self-driving cars and so on. It is premature too pursue it now. The biggest problems that we face is the humans. Humans are incredibly complicated and evil in many cases and lazy, greedy, violent and so actually reverse engineering all of that to figure out what it is we really want on our good days is a really difficult problem. And I think in the process of doing this we will actually learn a little bit more about how we should be, right? We'll not just be that way on our best days but maybe be that way more often. Thank you.

>> NICHOLAS DAVIS: Thank you very much Stuart. As we bring this speaking section to a close and Anja comes up to bring the panel on, one question for you Stuart, because we have seen an array of alignment here in terms of the importance of bringing the human being in to the discussion, but we have also seen some **tengs** emerge then between Manuela, desire to bring mobility in to robots and the concern you are expressing here before we do this we think to think provably and socially about the negative consequences of experimenting in this area. Is that a false

tense we are seeing here? A way to resolve that we can have our cake and eat it, too.

>> STUART RUSSELL: It is a good question. Up to now we haven't had to worry because AI is way too stupid to do anything. In some of our other talks I talk about domestic robot, there is nothing in the fridge and it cooks the cat. None of our **ro** bot could catch a cat, skin a cat and none of them know how to cook. We have been pure **u**ing a definition and approach to AI that only works when AI is stupid. And when AI starts to become sufficiently capable to have a real impact then it is that definition stops working and we have seen this already in computerized trading systems which have crashed the markets on several occasions and almost crashed the world economy in the process. And so once you get systems that can have an impact on a global scale, you have to think differently. I just want to say one thing why is it human centric. ? So Nick has a story of AI gone because it makes paperclips. It turns the **swir** world in to paperclips. If you are FipboxiDans which is a bacteria eats iron you are thrilled with a world covered in paperclips. We are the ones who are supposed to benefit and not the bacteria eating paperclips. It **neft** inevitable that the human will become the mathematical framework of AI.

>> NICHOLAS DAVIS: Thank you. In the ideal world we would would ask the speakers to come and take this forward. We have a panel that is Anja is going to moderate on the same topic. With that please can you put your hands together and thank fantastic speakers. Thank you very much.

(Applause.)

>> ANJA KASPERSEN: While everyone is getting seated I would just ask for indulgence to go over time. We started 15 minutes late. I apologize for that. We are trending on Twitter. We are No. 1 on Twitter. Thanks to all the the excellent twitter feeds and updates from the participants here. Quickly I will come back to each one of these as we proceed. So we have here Chris Fabian, with UNICEF. We have Lan Xue from the school of public policy and management at Tsinghua University in China. We have Peggy Hicks from the Office of Human Rights commissioner and we have Chaesub Lee who is the director of ITU standardization bureau and part of the organizing Committee that has brought all of us together. A big thank to Mr. Lee and his staff. We have Sam Moliics who is with the Zucker berg Foundation and at the end here we have Paul Bunje. With no further ado I will throw out couple of questions. We have been very rich discussion so far and so this panel is not really meant to bring up brand new issues though I hope there will be some holy cow moments but to reflect what has been shared with us by the distinguished speakers. To give some framing discussions, I hope you can think

of some interesting questions to pose from the audience when we come back to you. So let me start with the application side of this. So Christopher, you work with the UNICEF innovation unit and the ventures team which invests in a lot of these technologies which you call frontier technologies. UNICEF being an UN organization, working similar to my organization in very contested environment. What are some of the experiences that you have made applying these kind of frontier technologies in your situations? How do you deal with data integrity, the forensics of data being used in your technology, and what are some of the hurdles of how you overcome them. So Christopher.

>> CHRIS FABIAN: Hi guys. We're almost there. It is almost the end. Do me a favor before you answer your questions, raise both your hands for a second and just do this with your fingers. Doesn't that feel good? All the blood is going down. Thank you. All right.

Getting us through this last panel. Thank you to the partners for this event. You can find more on the UNICEF innovate twitter account and my Twitter account. Four things I want to raise and 30 seconds each. We have set up the **firgs** venture capital fund in the UN. We realize that with emerging tech **teks** we have to be able to make small early stage investments in open source company and pool their technology together and start to understand where the world is going. We have invested in **droens** and block chain and I would like to talk more about that on Thursday morning. Don't tell the ITU guys. The second thing I got like one laugh from that. There we go. Second thing who knows what this is? Raise your hand. One person. Two people, three people. Four people. This is is an muac. Mid upper ARM band. Is monitors the nutrition status of a kid. If you have a healthy fat little baby your band is green. If you are moderately malnourished it is yellow. If you are kid that's suffering from malnutrition your ARM looks like this. This is a valuable tool for famine response. We are looking at investing in companies that are using the same thing using face recognition. This is simple Artificial Intelligence but it is something that is important and affect the lives of hundreds of millions of children. How this works within the UN I ask to you Google a few things or use whatever search engines are. One is principles of innovations. These principles guide the basic ethical approach to new technology. Second the UN innovation network is body that ITU WFP, UNHCHR. The guiding that network have been important to us in bringing the community together and third products like the magic box. To give their data in to a data collective and use them for outputs allow us to work with groups like the Global Pulse and others to understand the world better. And finally partnerships like the partnership for AI for

good allows us to have that overall view of what's happening. So that's just a very quick set of things that we have seen over the last few years work as we understand and explore new technologies or don't understand them and happy to talk more about that in the question and answer.

>> ANJA KASPERSEN: Thank you for that and we will come back to some of the dilemmas that you didn't touch on. Sam I know you come from there data, more data scientist background. We talked and was mentioned by several of the speakers this open source as a platform for innovation. But although as Cindy was mentioning also as a tool to Democratize destruction. If it is acquired or used by people with malicious intent. From a data consumer experience side do you think around open source and ways of moving forward to ensure safety and how we developed AI and related technologies that AI infuse?

>> Sam: Absolutely. So to answer your question I will tell you a little bit about what we are doing and then I'll get to that. So the first thing that I want to say is that at METI we think that AI is part of the solution. In partnership with humans. In areas or especially looking at areas and activities that are really at the root of how we make innovations that underpin human prosperity and human potential and these are smaller markets from a revenue perspective but often have an outsized impact on humanity and science and education are two of these. And in fact, in many ways underpins solutions to the 23 SDGs that we are talking about. So within the sciences the speed of science in many ways is limited by two main problems which we think that AI is well positioned to help with. So the first is basically predicated on this notion that science is global distributed and decentralized activity amongst 10 million people in several hundred thousand labs and it is hard to understand what everyone is working on. Scientific awareness is a key challenge and the second one it is extremely difficult to **comprehend** everything that is now. Even within subfields and to reason across that and understand it is more than one human can do or a human brain can do. And so in metawithin chance Zucker berg we have been thinking about how AI can be used to solve or the use of AI to power scientific knowledge production itself and we are in the process of making available a global scientific awareness system that can accelerate advances in sciences and make it available for free, open to scientist, researchers funders, policymakers, et cetera, and one of the reasons why we think is this is an important thing to do scientific advances and discoveries themselves are recursive. And science compounds itself and this is nonlinear thing which is important. Building on themes that we heard earlier this morning. A great example of this is the discovery of crisper which is method that's un

locking a variety of new stem cells and the cycle continues and accelerates that we have seen for a century. From the perspective of free and open source we have been working with the -- operate on all scientific content, including closed access content, there is tremendous way to do that based on this idea that it is important for everyone to know what's happening in the sciences and our tool as all tools in that operate in this way is available for free and should be made available for free to individuals to institutions, to teams and to enable everyone. So that's basically how we think.

>> ANJA KASPERSEN: Thank you for that. So the last person I will ask more on the application is Paul Bunje, one of the co-organizers of this event. You come from a very strong background of global conservation and environment. We are pretty critical juncture in terms of future commitment to environment. How do you see AI being used in this context?

>> PAUL BUNJE: . Thank you very much. It is a great question. I also greatly appreciate the pronunciation of my name. As an American we go coarsely with Bunge: There is a [nom](#) of great examples even on the wheel behind of ways that AI can apply to a number of Sustainable Development Goals and those in the environmental space in particular. It was mentioned earlier that right now the UN is also hosting a conference on oceans. Simple machine learn willing comes from a program called global fishing watch. One of the things thissy have been able to do is use to utilize the data in fishing vessels and apply machine learning algorithms to identify where those vessels but what kind of activities they are engaged in. Those data have become so successful that there is even one case where illegal fisher was fishing in a protected area of Palu and was successfully prosecuted as a result of the data that came out of this. Even though they then offloaded those fish to another vessel out, another ship outside of the area. It is a very simple thing to have been applied but it is transformative. Because efforts very often conservation or the environment are one example but this is broadly true of a number of Sustainable Development Goals where the resources to fight the problem are often quite lacking. We might be using these very linear approaches of catching someone and then enforcing the law in some ways if the law does exist. But those don't scale very well and though the malicious Actors have the ability to circumvent a lot of these efforts quite [waud](#) broadly. In wildlife trafficking, increasingly on the ground where you can tag animals, for example, or utilize microphones to identify where it is that illegal logging activity might be happening or poaching activities based on shotguns fired those are starting to be utilized to predict now where the legal activity might lead and

then particularly excited by some folks that are looking at using that in black market transaction financial transaction, very often illegal wildlife trade, using the same routes of laundering money as to human trafficking or narcotics and others. There is a lot of examples where tools used perhaps in larger market driven senses are being able to apply to things like conservation.

>> ANJA KASPERSEN: Thank you for that. It relates to Cindy' s remark around the crime. If you watch X men, the ultimate peace keeper. You do the same thing we are trying to preserve our planet. I am not sure if we will be sticking around or the optimization of goals the only way to preserve is get rid of the rest of us. Going to the bit of the governance kind of right perspectives, so Peggy Hicks of the Office of Human Rights commissioner, and also specialize in Human Rights. In this hyper connected world around dealing with rights and data and who owns the data and fore~~ren~~ sicks of data that goes in to particular cognitive technologies do we need digital Avatar to help Is protect ourselves or to help you protect us or to have a new set of rules?

>> PEGGY HICKS: Thanks very much and happy to be here. I have heard a number of things that then to answer that question. I want to harken back to a comment that Robert made, one of the big challenges here is navigating a course between the risk of doing and the risk of not doing and for me when you ask about a digital Avatar how we bring Human Rights or ethical values in it is about how we figure out those challenges because we know we need technology. AI can be a huge resource as we heard in ~~compat~~ing the world' es problems and brings with it a host of challenges. I would like to make three points on that. Human Rights is really a useful tool in helping us navigate that space. We are not free floating in a world without any values or principles that have already been defined. We have worked on this project for many years and we have ideas and the universal de-Clairation of Human Rights which was referred to earlier founds us and grounds us in principles like dignity and nondiscrimination that can be the watch words. We also have really extensive experience that I think is uniquely transferrable to some of the challenges we will face in looking at the risks of doing with Artificial Intelligence and so Robert also spoke about soft law and how we develop it. So we have experience in things like the guiding principles in -- we are not sure what the space is bringing in the existing system of Hugh manrights and developing it in to standards and hard law that can help us navigate a more difficult space. We experience in both normative development and looking at how we monitor these spaces. One of the things we heard from Brad Smith at

Microsoft is there is a need for an international or independent place that will monitor and look at some of these questions. We can't leave it all to business or to government. We need to find a separate space to do it and I do think some of the analogies with the Human Rights sphere can be helpful tools. I wasn't to focus on, thought about as mind the gap. In the British express. It is not a gap. It is a **kasism**. Which is the focus today. And so when we talk about applications and what we really need I think one of the main things I would like all of you to be thinking about in the next two days how do we close those gap and bring those two spear spheres together in an effective way. Simple things like educating more people in both fields is a key step in doing and then finally one of the principles of Human Rights that have to be brought in to this conversation is participation. And we know that the challenges of that in a field that is so complex that it is difficult to bring people along to even understand what we are doing much less support it and Robert talked about the need for demonstration projects. It is more than just building trust. People fear what they don't understand and getting people to understand this is going to be very, very hard. So we need to find ways to bring Civil Society, and not just Civil Society focused in this particular space but a broader range of people in to the conversation.

>> ANJA KASPERSEN: Thank you. Important points. So Mr. Lee, ITU established in 1860 to govern the electric traffic between states, Internet, mobile technologies, where do we go from here? And just want to refer to a recent pewe research report that was describing a situation looking at the U.S. audience but looking more globally where people are more confident in industries protecting us from some of these risks that Digital Age and new technologies bring on than governments and this can being a Government organization and we need this kind of multi-stakeholder Government rooted as Peggy was saying both standards but also hard law. Where do we go from here.

>> CHAESUB LEE: Thank you very much. On behalf of my staff I welcome. I lost my chance because of another event just across the street. We had IoT week is another big event. IoT is one of our event with the support of this AI. Let me challenge this, technology development is quite a controls. Let me say this AI, AI is nothing new. AI is quite old tough. AI is coming from many individual parts. We call it maybe in our language is protocols. Protocols they engage of this development of AI. Development of this AI we have a certain common elements like something data. So we have a situation of this vertical development and some of the common areas called horizon elements. This is a situation. It is the same as what we developed now UN systems, we develop our industries, verticals

but we have more and more commonalities called the name of this convergence, whatever use of these ICTs. That's very good benefits but certain some challenges for us. So as UN agency of ITU we -- we look at this very carefully and we recognize this ICTs really common for that. And when we recognize the ICT also innovated by AI but okay, that's -- true innovator of this ICT we have to understand what is the impact of AI. So this -- right of the way. Listening of this sessions I'm thinking about this we may single out futures how we can go forward. And try to analyze of this front classifications, like associate issues, like privacy issues or Human Rights. Another limit might be business perspectives, we need some business -- we have to facilitate of this business, creation of this new business, new markets, and final element might be technical elements. What kinds of technologies always impact of this outline. Maybe some different ways how we can move forward and how we can collaborate. So this is I believe this is a challenging subject for us of this AI Summit because ITU is a part of this ICT. But other UN agencies other stakeholders you have your own views. For example, we may think about this, the concepts of this essence by design, human centered. Someone highlighted human centered approaches and we need to talk about associate subject. We tried many times by design concepts for announcing security and privacy but unfortunately we are a little bit late. The security is already gone. The privacy is quite a difficult to say force of this by design but now it is a good challenge again, ethics by design. Why not? Because this AI just start. This event is very important. All stakeholders get together now. So it is good time to strengthen this ethics by design. Human centered approaches of this business. Some of -- business and technologies we have to think of how we can strengthen of this verticals like continue of this business development. But also how we can show of this horizon elements like data sharing. Data we know is AI should be produced result. We are based on input data. So if data is -- definitely -- so we have to something, to make clean of this data to ensuring the data while we are protecting the privacy as well. This kind of concept will be necessary I believe. And also we have to strengthen of this many peoples like Cs, collaboration, coordination, cooperations with many stakeholders. So we are ITU is a part of multi-stakeholder models. We are delighted to work with you and looking to the future. And we are expecting this next two days we can find some ways, what is directions. How we can do. If we can find some ways it might be very desirable event and very productive event.

>> ANJA KASPERSEN: Thank you Mr. Lee. So key role played by ITU here. So continuing on the design thinking which and design kind of relating to governance professor Lan Xue worked

in many different processes trying to bring to the forealigning signs with design and with governance and actually your Ph.D. was on engineering and public policy which is, you know, you are silo buster on stage which is great. I would love to hear your views on this. How do we bring the design in the to governance?

>> LAN XUE: I think a lot of colleagues have talk abouted various good wishes how of to harness the AI technology for the good of human beings and how to minimize the risks but I think there are three major challenges that we have to overcome. The first one is what we call the regime complex issue. AI is so complicated that there are many so-called rejems that's related to AI. We have ITU and also many other international organizations that are related to many aspects. They are not hierarchical. So they each have some bearings to the AI to the advantage, to the risks but they cannot actually -- dominate one regime cannot dominate the other. How do you coordinate with these rejooems. That's the first challenge. The second issue is what I call the cycle gap. Because of the complexity of these regimes. In order to find in a policy to guide the development of the technology the process is going to take much longer time than the technologies cycle. So in other words, technology moves much faster than the governance cycle. So I think that gap actually will not make, you know -- so actually the gap of the two make it is very difficult to have effective governance system. And the third is outcome the first two is fragmentation of efforts. So you have a lot of people a lot of institutions trying to work out different things but at the same time it is not going to make, you know, huge impact on the major trend of the technology. So how do you address those two -- those three obstacles? I think the first one I think that we can think about is the -- I think we can learn from the climate change. Another sort of major global challenge we are facing. Think the IPCC mechanism is one that we can learn from. In other words, creating sort of knowledge, global knowledge base doing kind of analysis, comprehensive analysis so that after providing a knowledge platform for policy for action.

So I think, you know, so one proposal I would make here, there is probably a need for global knowledge alliance for AI that at least we have a clear sense of where the technology is. What the potential impact. Both positive and negative. So that actually, you know, that would be serve as a guidance for the industry and also for the Government. And the second thing that we can also think about is to think about sort of taking adaptive governance approach. I think when we designed the policy we always wanted to be comprehensive and wanted to be, you know, covering everything and also being very effective but at the same time I mean given the complexity and given the uncertainty related to

the technology I think actually we probably need to think about adaptive governance which has been used in many other areas to allow the some ambiguity to allow some sort of uncertainty in designing where it is policy and guiding the development of the technology. So that actually designing a learning process, so that in a way you can move from one step to the other and then learning and adjusting the policy that we can move from there.

>> ANJA KASPERSEN: Very profound ideas. I am going to open up audience. Please introduce yourself and keep your questions succinct or comments to the point. I will start over here.

>> Thank you so much. My name is Dr. David Mebnu. I am working with a startup famed Alfred and psychiatry machine learning. Two questions. One is a serious one and one is thought provoking one. We heard a lot today about the potential for autonomous weapons system and the other way that AI could end up directly killing people. I would like to ask something about already killing peep is targeted advertising through AI. We heard about the challenges of chronic disease. Chronic disease is often informed by lifestyle choices that are also often exacerbated by smoking, alcohol, poor food choices much of which is driven by advertising in which now has this added weapon of Artificial Intelligence. So are we going to be answering a place where AI used by companies is going to be in competition with AI used by social good organizations like the UN one doing education, one doing advertising and how do you look at corporations and have them resolve that conflict while also having an incentive to profit from targeted advertising, for example, selling alcohol to somebody you can tell from the press and social media. Second question we have heard a lot about human centeredness. But what about AI rights when they get complex enough to deserve them. If they do.

>> ANJA KASPERSEN: Over here there was a lady in the back. Yes.

>> Yep. Thank you very much my name is Kristy nan. I am at Chair of European group. My -- thank you very much for these talks. Artificial Intelligence can do a lot about understand can problems and analyzing them doing diagnosis of every kind. But is there a power gap not only a cycle gap but a power gap in solving the problems? So what's about diagnosing a disease but not being able to treat them. That's my question to the panel. So Artificial Intelligence seems to be very powerful but not in solving the problems so far. Sorry.

>> ANJA KASPERSEN: (Off microphone).

>> Yes. Hello. My name is altan from the school of hygiene in the UK and I am working on chronic noncommunicable disease analytics but my question related to my previous life where advisory to former Prime Minister David Cameron and I have a big

question and it is unfair to ask the panel at the end of the first day but I wonder if we can land the question and it can be tracked during the course of the next two days. How does this dialogue fit in to current experiences around global trade, patent, and commercial legal practices? And I what I am trying to bring out here if we are saying there is clear and very specific social good related to this technology, which in fact, is a commodity, consists of products and service as continue highly likely it is going to fall in to our frameworks for trade. If there is a strong social value being place around this agenda. How do you keep this social commitment and do it in a way that's effective in terms of trade processes. Think about frontier markets. Frontier markets are places where people are looking to move in to. And also Developing Countries and, of course, we know from our experiences that that's also about shaping markets. And I think the question is essentially that, how do we dialogue with global tariffs and trades around this. Should AI be in fact, a special economic zone which is free of trade and tariffs and taxes. That's my question and could this question be put to the WTO and other multilateral institutions with that responsibility. Thank you.

>> ANJA KASPERSEN: As you said a big question and is can simmer and mature over the next two days. First a gentleman all the way in the back and then you.

>> Thank you. Richard Hill Association for property Internet governance. I'm a Civil Society activist although I wear camouflage and I was very pleased because at this conference just in the first day I have heard every single theme that I think is important. Both in general and specifically but I wanted to drill down on one specific, which Ms. Hicks you mentioned one of the initiatives from Microsoft in terms of cyber attacks but perhaps you could comment on another which is the call for Geneva digital Convention whereby states would voluntarily agree to refrain from certain activities. Thank you.

>> I am Ragish from Germany. I'm first of all very much thankful and and I am very much pleased to attend this conference and I only think that this conference will grow bigger and bigger in the coming years and centuries. I would like to say that quote Albert Einstein, he said technology is not helping us solve the problems in the world because we have not solved the problem of international relations. So he meant United Nations should be further empowered and we should come out of nationalism. So from World War I we move from legal of nations and World War II we moved to United Nations. I wish we don't have any further ka that stro fibbing event for world Government. One common plan. My question is how are we going to make sure all the regulations are going to be followed towards

the world because we have so much partnership on AI but who is going to and how are we going to make sure that these regulations are followed throughout the world and I potentially feel empowering United Nations is only the way. And further such discussions should be made in future.

>> ANJA KASPERSEN: Thank you. You should go to your Yoshua's panel on AI for prosperity tomorrow which we want interesting discussion around this. And then was two more questions and then I'm closing for now to leave it back to the panel. Sorry it was -- yes, you sir and then here. Yes.

>> We spoke about exponential growth and this links very well with AI. But exponential growth means that whoever come first using the AI can get such an incredible success that it destroys everyone else. If I start selling ice creams and I have a very powerful AI system to advertise my ice cream I can become the best ice cream seller in the world. But any other guy selling ice cream will be out of the business. I'm concerned about the people being left behind by the exponential effect of AI and the fact that someone will be incredibly successful. Of course, is the lead story that will come to the newspaper. But all the millions that be will left behind probably should be remembered, too by us. My name is Massimo. I am inlek [ral](#) property lawyer from Italy.

>> ANJA KASPERSEN: Thank you. Last question here.

>> So yeah. Thomas I am from the Fraunhofer from Berlin and from the Technical University in Berlin. So my question is how do you deal with the fact that a human is a human and a machine is is a machine. That's obvious. But if you think about it, a machine at its core can never be guaranteed to respect human values because it doesn't have the neurological necessities, right? So everything can be hacked and, of course, humans turn on each other but that percentage is small and that's why society is stable. So if you think about it, then one step further when some of the great talks we heard before when AI becomes smart enough to matter, don't we then need a security check on all AI systems all the time? So when you enter airplane and you go through security because no human can be trusted for some reason and no AI system can be trusted. There needs to be security check for self-driving car. And we also need to have address another fact which is the fact that networks and malicious AI system can have once it is networked, and maliciously hacked it could, you know, have a mass destructive effect although it is designed a completely different way. These are two questions and I am sure there will be lots of other questions and which organization will be regulating AI systems in the future and how do the panelists see that.

>> ANJA KASPERSEN: Final question. So given time we have

left what I will do is I will leave to the panel you have all the questions and let me just quickly summarize the question was around targeted advertising and the danger of it. It was about AI power gaps and solving versus treating. It was dealing with diseases and, you know, thinking around practices and social good. And it was AI as a commodity. AI is in a trade relation and data is already part of the trade discussions, will the refined version when it turns in to algorithm or cognitive systems be part of that equation as well. And then it was the comment on the call for a digital Geneva Convention which is thing attribution of cyberspace and how do we make the discussion truly mobile. And make it truly global and what is right platform and it has to be the UN which was the suggestion and then the last question if I get this correct now was -- second-to-last question was exponential growth and ensure we are all on the same playing field and have the cognitive augmented capabilities to not put all other ones aside and the last one to be more of the adversarial question and how do we vet and validate and verify how the AI has been coded and come to determine our lives or **ke** decision making process. Those are **proet** profound questions that spans over wide area. I will just give the word for to each one of you in reverse order. And would ask you also think about the how and think about what you want to leave all the participants with in terms of thinking about how to prepare for tomorrow's workshops. We will dive in to some these of these questions. So we will start with you.

>> ZUE: I think there is several issues related to the go governing of potential risks technology to go through the market and being widely used and then we address the risks and the problems at the end eye cycle. I think with AI technology I think what all people are raised in terms of concern we need to start at the **giP** beginning. We need to think about what system are we going to adopt to address the risk issue. And so that actually we can really embed the system and the principles in to the designing and the -- of the system. In to the application. I think that's probably the principle I would follow.

>> ANJA KASPERSEN: Mr. Lee.

>> CHAESUB LEE: Yeah. Those questions are really profound questions and not easy to answer. One thing is we assume AI implemented completed just tomorrow. That's not going to happen. AI is -- continuously grown up. And now it looks very close to us. But it is still far away to complete, I don't know when will complete but AI continuously grown up. Even our society is grown up. Since we use this mobile phone, how many years now we use of this Smartphone? Around 15 years, 10 years.

>> ANJA KASPERSEN: Nine.

>> CHAESUB LEE: We have changed many parts but still we

have the change in process. So we can raise of these questions for the target goals, target goal times, raise of the question in this today. Might be something struggle ourselves. Rather than I want to say of this as a step by step, like someone talking of this exponential growth, we may think of that -- that possibility. But we believe I believe it is old -- during that development of this AI process, AI enabled ice cream, AI enabled ice cream controls, our society, our business our communities they develop their own ways. So will be relatively shorter gaps. But that question would be very important for us because we are now want more fair human centric means. We look at this fields, there is a 17 goals of this. We are all human beings now want to realize implement of this 17 goals. Which means that ensure this business about profit, that's also important but it is time to think about how to meet of this SDG. Business insurance is very important. I don't know regulator of this but it is time to get different way of balance. So winner take all, we can continue these ways or we can think about how we can show the benefits. This is another element. This is good time to think about how we can really share of this benefits to make people have a more sharing of this human beings, well beings how we can make it happen. This is important part. And in this regard some of regulatory aspects, so can control of this, I believe each verticals, especially our colleagues from the UN organizations, each vertical has some UN organizations have responsible for each vertical areas and we have certain responsibility to address that. Some of our groups, I believe ITU in terms of this ICT, we have certain roles of this horizon areas. We have our responsibility to address this horizontal part. If we can have some good agreement of such directions we have a more deep investigations how we can practically implement of that areas. This is my view.

>> ANJA KASPERSEN: Thank you. So Peggy.

>> PEGGY HICKS: Thanks. I will take up the question on Geneva digital Convention. The UN Human Rights office doesn't have a position. Important things to I think the analogy that Brad Smith has brought to this is that this is like a weapon in a sense and we have to deal with it as seriously as we do deal with other weapons. Is it a really important way to garner attention to ab and to think about how we take up this challenge. It is also important if you believe there needs to be a Convention to realize that we can't wait for a Convention. Get more immediate action because, you know, this isn't a process that can wait five, ten years before we can have a full-fledged Convention. That's why this question of a monitoring facility, inform more legal or normative process is an interesting starting point for the conversation. Could I also just go back

to the question that started us off about advertising and social good versus the those that are using corporate promoting corporate goals in a different way. I think it is really important question that sort of should underline some of the conversations in the next few days. The reality is we are talking about how AI can contribute to discrimination or inequality globally but we already live in a very unequal world. So we are starting not from a level platform or. We are starting with huge deficits as I know na Salil Shetty talked about efrl yerl today and we are to figure how can we not only invest in the side of trying to make sure we don't make it worse but using technology to improve and address those gaps. And to me that comes down a fundamental question of investment. We will be as societies willing to see this chance to meet it and to invest in using Artificial Intelligence and other technology for social good and initiatives like that one just described by the Zuckerberg Foundation are critical starting point but it is going to have to be much broader and much deeper.

>> ANJA KASPERSEN: All those are in very vul ner able contacts. Paul.

>> PAUL BUNJE: Thanks. I think I'm going to build a bit on what folks are saying and I picked up in both those questions on throughout to youed on a couple of themes that are relevant and how we codify those values in a program for developing AI and AI for good in particular. And the other being exponentials. Something like targeted advertising or the use of Artificial Intelligence to more effectively make our weapons kill people may also be considered human values. Not all values that human maintain are positive in nature. Of the point about exponentials and this notion of being able to digitize, puts the potential in it a small group of individuals sh the power that one took large linear institutions. You are decoupled with tech teks like AI. Be they ones that we all might share or ones that vary widely. So what does that mean? Where do we go when you asked how do we deal with this. Let's pick up on this notion of Artificial Intelligence, like any technology can can used for good or for ill and we often look to specific institutions to represent some sort of codified social norms or social contract and then regulate and to place what we want to see happen. But I think also picking up on the idea of how institutions work and the theme of this panel collaboration we need to recognize that the world is changing so dramatically that things like states are undermined by the very technologies we are promote can. Have the ability to use utilize these things across state boundaries and implement whatever use they might value notion Ali at that given time. So we need to think about how it is not we just use our institutions to regulate but we have an expression at XPRIZE

that you get what you incentivize. We need to start thinking about the other tools carrots in the mix that will allow us to demonstrate what are more likely to be shared human values and `ensint tooif z uz` the. And I think that's where this -- two sort of themes that are coming out of this. One why leave -- we exist right now a very opportune time. We heard about the growth exponential growth and heard about some similar shared principles or at least notional general ideas that AI can be used for good but also heard that it is not there yet. AI is still not very well developed but puts in a good position to think about this roadmap for the future and you heard Mr. Katsumi Emura talking about as well as professor Russell how we think about a vision of the future and building out a meaning plan to achieve it and over the next couple of days we have the opportunity to put some of those very beginning steps on to roadmap that we can execute against. That's a rare thing in human history to have that ability of foresight and ability to plan and we heard about this need for collaboration and there are lot more conversations like this to be had. Who the participants are, what the sort of rules of the road that might look like. There is some notions here but it is not going to be easy to come up with I think the sort of firm next steps that are happening. It is only going to happen through further dialogue.

>> ANJA KASPERSEN: In AI `stran` sends and Europe borders permeates everything we do our mind set and responses need to correlate. What you --

>> PAUL BUNJE: We need to build our institutions and our society to recognize that very fact.

>> ANJA KASPERSEN: It goes to your point who is then the right place to do so. So this is a starting point also to discuss that. Sam.

>> Sam: Thanks so much. So I think what I would say cutting across a couple of the questions is that there is a big new tool available and Governments not-for-profits philanthropies are in sort of an unprecedented position to turn their dollars in to social good. Cases where you can ways basically find a problem that's an information challenge or translated in to an information challenge where to drive for it in AI project and then make that available to everyone through the Web. And I think, you know, that could be fighting dangerous advertising or it could be new health challenges and think this is a really big deal. And then the second thing that I would say is related to the question over here about sort of the `asem mi tri kal` advantage for first movers with AI in a market a lot of new advances that `koom` from machine learning models that are reduced to practice become commoditized and a lot of the

advantage comes through you unique datasets to train.

>> ANJA KASPERSEN: Thank you. I am waiting for your next stint of mobilizing the audience as well before the drinks. So you can close off here.

>> CHRIS FABIAN: Never give away the Moderator mic. Thank you. So I started sort of being funny but I think we -- I would like to end on something quite serious. And hopefully this is something that underscores the discussions in the next two days. It is he had easy to think about all all the great stuff we have haired today and the all the technology that we have seen, and think about the wonderful future that awaits many of us and that's true for many of us. If we are wealthy and if we live in urban areas we are fine. But the world is entering a time of great darkness and if we are thinking about the next 30 years if we listen to what Jack Ma said the next 30 years hold more sorrow and main for humanity and they do joy. Answer these seven questions with new numbers and facts and we will end this. There are 55 million kids who are on the move today because of war and violence. 55 million is a number that I cannot make anybody understand. These are kids who don't have a state or sovereign Government or institution to fall back on. There are 500 million kids who are at risk not getting the right nutrition because they do not have access to the right food because they don't have water, not enough anymore other they have too much water and these are because of climate change. If you are a kid who is moderately or you a cutely malnourished your brain does not develop. And that create as asemi trick world. 3 I talk recently to head of large airline company and he shut down his pilot training academy. They are not training human pilots any more? Why? They have enough humans to last until they can do the piloting. They will felt most acutely in poor places so you have a country like Malawi where you got 18 million people and of them 50% of them are kids. 10 million kids. In the next 15 or 20 years population grows it will gre to 40 million and that's 25 million under 20 years old and there are the not enough jobs and how do you answer them? There is very little that we can do. We heard about a post sovereign world and there are more people on Facebook than people in China. Those services that platforms are providing sovereign services cash transfer education, identity, and we as the UN do not know how to negotiate with them do not know how to work with them and how to set policy with them. It is a new world and very scarey world. We as UNICEF is have mandate and mission to advocate for the rights of most vulnerable children and the 55 million kids who are on move don't have government that we can relate to. So I think that as we look to the next five years or ten years and I challenge everyone to think about this for the next few days. We

can think of AI solving the problem you who do we solve the problems for all of us. With Facebook and Google and I beshgs M and these other big tech partners who want to make money but also want to do good. How do we create a world of fairness and understand our position. These are the sort of questions that I think I would actually raise as secondary questions to some of the ones that we heard and I very much look forward to discussing them with all of you in the next two days. Thank you.  
(Applause.)

>> ANJA KASPERSEN: Thank you so much for those final points. Just to add my organization also work with people on the move and in very vul ner inlible context. Other forms of serious violence or crime. And we in addition to what you are saying about the people suffering there is also reality of that increasing number of these are connected and they generate data and how do we deal with the protection of data when you are at your most vulnerable. Every single one of these kids grove growing and they are going to apply for next bank loan and data may have followed them for 15 years. So there is a lot of issues that comes out of this ifrn creasing number of vulnerabilities, different levels of literacy understanding and also being able to say no, you can't have my data. You are standing there in a refugee camp and you need to give away your data to get the food you need. Certainly to what will eventually feed an AI in these contexts. So I want to thank the speakers, the panelists, and Nick you had to leave to catch a plane. The organizers for putting this on. And just kind of again encourage everyone to when you go back home you sleep like really let this discussion today simmer and come back prepared to tomorrow. And let's really start working on the how. We have two more days to figure out the how and who governs and how adapted we can be. I'm really looking forward to that. Just FYI housekeeping things outside is going to be a reception as an I look forward to engaging with people there and there is going to be drinks and some refreshments and other things, right? Fred?

>> (Off microphone) drinks reception will be at the cafeteria in Montbrilliant. So you have to go up two levels of stairs and down a narrow hallway.

>> ANJA KASPERSEN: Thank you so much everyone and sorry for keeping you late.

(Applause.)

(Session concluded at 6:27 p.m. CET)

\*\*\*

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.  
\*\*\*