The National Academies of SCIENCES • ENGINEERING • MEDICINE

Board on Science, Technology, and Economic Policy Committee on Science, Technology, and Law

Workshop on Robotics and Artificial Intelligence: Policy Implications for the Next Decade

December 12, 2016

UNEDITED TRANSCRIPT

This is a transcript of a proceedings of a workshop on Robotics and Artificial Intelligence: Policy Implications for the Next Decade, prepared by CASET Associates, Ltd., and is not an official report of the National Academies of Sciences, Engineering, and Medicine. Opinions, statements, and presentation materials included in the transcript are solely those of the individual participants at the workshop and are not necessarily endorsed or verified as accurate by the National Academies.

Planning Committee for Economic, Legal, and Regulatory Implications of Emerging Technologies: A Workshop Series

Ryan Calo, University of Washington (chair) E. William Colglazier, American Association for the Advancement of Science Mary Cummings, Duke University Henry Greely, Stanford University Simon Johnson, Massachusetts Institute of Technology Milind Tambe, University of Southern California

<u>Staff</u>

Gail Cohen, Director, Board on Science, Technology, and Economic Policy Anne-Marie Mazza, Director, Committee on Science, Technology, and Law David Dierksheide, Program Officer, STEP Board Karolina Konarzewska, Program Coordinator, CSTL Frederic Lestina, Senior Program Assistant, STEP Board Erik Saari, Senior Program Assistant, STEP Board

The National Academies of SCIENCES • ENGINEERING • MEDICINE

Robotics and Artificial Intelligence: Policy Implications for the Next Decade

December 12, 2016

Room 125 National Academy of Sciences 2101 Constitution Avenue, NW Washington, DC

Agenda

- 8:30 9:00 **Welcome and Overview** *Ryan Calo, University of Washington, Planning Committee Chair*
- 9:00 9:45 **Keynote Address: The Opportunities and Challenges of AI and Robotics** Speaker: Jason Furman, Council of Economic Advisors
- 9:45–11:15 **Panel 1: Economic Impacts: implications and opportunities for growth,** economic mobility, and inequality

Moderator:	Simon Johnson, MIT
Speakers:	Caroline Atkinson, Google
	Guru Banavar, IBM
	Martin Ford, author and futurist
	Milind Tambe, University of Southern California

11:15 – 11:30 Coffee break

11:30 – 12:45 Panel 2: Ethical, Legal, and Regulatory Considerations: considerations of privacy, discrimination, and international norms

Moderator:	Hank Greely, Stanford University
Speakers:	Missy Cummings, Duke University
	Robin Feldman, UC Hastings College of the Law
	Deirdre Mulligan, UC Berkeley
	Laurel Riek, UC San Diego

12:45 – 1:30 Lunch

1:30 – 2:45 **Panel 3: Security Implications: implications for public safety and national security**

	Moderator: Speakers:	Ryan Calo, University of Washington Douglas Maughan, Homeland Security Advanced Research Projects Agency Stuart Russell, University of California, Berkeley Gaurav Sukhatme, University of Southern California	
2:45 - 3:00	Coffee Break		
3:00 - 4:30	Panel 4: International Session: policies around the world		
	Moderator: Speakers:	John Zysman, University of California, Berkeley Robin Mishra, S&T advisor, German Embassy Kenji Kushida, Stanford University Dan Breznitz, University of Toronto	
4:30 - 5:30	Keynote Addr Speaker:	ress: Powerful But Limited: A DARPA Perspective on AI Arati Prabhakar, Defense Advanced Research Projects Agency	
5:30 PM	Final Remark <i>Hank Greely, S</i>	s Stanford University	

5:45 PM Adjourn

Table of Contents

Welcome and Overview Ryan Calo, University of Washington	1
Keynote Address: The Opportunities and Challenges of AI and Robotics Jason Furman, Council of Economic Advisors	11
Panel 1: Economic Impacts: implications and opportunities for growth, economic mobility, and inequality Moderator: Simon Johnson, MIT	50
Panel 2: Ethical, Legal, and Regulatory Considerations: considerations of privacy, discrimination, and international norms Moderator: Hank Greely, Stanford University	138
Panel 3: Security Implications: implications for public safety and national security Moderator: Ryan Calo, University of Washington	194
Panel 4: International Session: policies around the world Moderator: John Zysman, University of California, Berkeley	249
Keynote Address: Powerful But Limited: A DARPA Perspective on AI Speaker: Arati Prabhakar, Defense Advanced Research Projects Agency	321
Final Remarks Hank Greely	363

P R O C E E D I N G S (8:30 a.m.)

Agenda Item: Welcome and Overview

MR. CALO: Welcome everyone. My name is Ryan Calo. I am a law professor at the University of Washington, where I also co-direct the interdisciplinary tech policy lab. On behalf of the National Academy of Sciences, I am delighted to welcome you here today to this day-long workshop on the policy implications of robotics and artificial intelligence.

Interest in robotics and artificial intelligence is very, very high today, but the technologies have been around for quite a long time. We have had robots in the United States for many decades. In fact, actually, as a law professor, I was interested to discover that they have already occasioned some really interesting legal disputes. I don't know if you guys are aware of some of them.

For example, as early as the 1950's, courts had to decide whether or not when you imported a robot from another country like from Japan - they had to figure out whether or not that was a doll for purposes of a tariff schedule. To figure out whether something was a doll, you had to figure out whether it represented something animate. There are these court decisions that go into great detail about what it is to be animate and the nature of life and

whether a robot is animate or not just to figure out what the tariff schedule. This happens - by the way, what the court decided in the first case was very interesting. It was that while robots are mechanical men and they represent something animate, a toy robot only represented a robot and therefore, is subject to a different tariff than a doll. This came up so often that, actually, eventually, the tariff schedule had to be revised and include robots as one of the categories. This is in the 1950's.

Another case is one of my favorites of some mischief that robots caused for the law. Do people in this room, especially those of you who have kids, know of Chuck E. Cheese? Do you know those restaurants? These are these terrible - they are wonderful, terrible restaurants. I am seeing a lot of heads shaking in understanding. So, at Chuck E. Cheese, for those who are not familiar, it is a kids' pizza place. They have these amazing animatronic kind of stuffed animal bands that suddenly sort of come alive and like play a song. Some clever tax authorities in Maryland said, well, goodness, if you are going to have these performances, then we should charge you a performance tax on food. They went in there and they tried to charge them because of a performance act. A court had to decide, for example, whether or not these animatronic robots were

performing. It went into all this detail about the nature of performance and how it required spontaneity and the like.

So, robots have been around for a while. They have already caused legal disputes. They have already caused policy problems for taxation and others. What is different today, I would argue, is that robots are a lot more versatile. This is largely due to changes in the costs of sensors and actuators and processors and the like. Costs have come down dramatically. Certain techniques have become commonplace like cloud robotics that is potentiating quite a lot more variety in what robots can do. As a consequence, of course, robots are leaving the factory floor and entering the mainstream. They are entering the air. They are entering our roads. They are entering our hospitals.

I think the same could be said of artificial intelligence, in general. That term was, of course, coined in the 1950's at my alma mater, Dartmouth College. The algorithms and techniques that underpin today's applications about which we are so excited have been - were developed, essentially, by the 1970's. They are being refined, but the basic techniques have been around for some time. What is different now is that we have a lot of data. We have cheap storage. We have really fast processing. I had occasion to help the White House put on a series - some of the people in this room were involved as well - in putting on a series about artificial intelligence and policy over the summer. In connection to that, I got to talk to quite a lot of AI pioneers like people in this room. What they said was they had the same algorithms then, but they would like try to do something where we would play a game of tic, tac, toe, and they would have to leave the computer on for a week in order to get the next move. It is basically the same sorts of applications we are able to do today, but it just took a very, very long time.

I also dug in and I was curious about whether or not the National Academies, itself, might have looked at this issue. Sure enough, in 1987, people were sitting not in this room necessarily, but in this building for a workshop that was called Army Robots and Artificial Intelligence and their Legal Implications. This is not something that is necessarily entirely new. Just because something isn't new, doesn't mean that it is not newly interesting. You also talk to people in artificial intelligence and they are genuinely surprised at the pace at which some of these things have happened. In

particular, voice and imagine recognition has been surprising to many people who are deep practitioners of artificial intelligence.

My favorite example is Skype Real-Time Translate. This is a system where you can talk to somebody in another language. It will detect your language and theirs. It will translate - it will recognize what you are saying in real-time. It will translate and put it at the bottom and you can have a conversation with somebody in another language. Some of the things that IBM Watson is doing, it is like magic. You usually need like a Tardis from Doctor Who or a Babblefish from Hitchhiker's Guide to the Galaxy or you need to be on Star Trek for some of these things to happen. It is not to say that just because these things have been around for a while that it is not a time to be really excited.

In my view, robotics and artificial intelligence will have serious repercussions for all aspects of society. I will just give you one example. How does the law react and how do policymakers react to systems that are able to set their own goals? We invited a guy named Peter Stone from University of Texas. I don't know if you guys know Peter's work at all. You know Peter's work, those of you in computer science. Peter wasn't available, but I want to talk a little bit about his work and what I think it - it is illustrative of some of the issues we may confront across all aspects of society.

What Peter does is he works on systems that have ultimate goals that are dictated, of course, by the programmer. In order to resolve those goals, what his students work on - what he works on is allowing the system to set its own sub-goals and advance the ultimate goal. Those things can be surprising.

For example, he tells this story about one day, he has this robot and the robot is trying to map out the building. It is trying to map out the University of Texas computer science building. It is supposed to like go around and explore and make a map of it. It is pretty interesting stuff. One day, his Ph.D. candidate – actually, one night, in the middle of the night, goes in because they forget something and they come across one of these robots. What it is doing is it is just basically going through a door and then turning around and going back and turning around and going back through the door again over and over again. All roboticists are used to these things, like, oh no, what is going on here? I shouldn't interfere with the experiment. There is obviously a glitch. We will look and see what it is is tomorrow.

So, when they go and they look and see what it is, it turns out what they system had figured out that if you are mapping a building and you are a robot, one of the most difficult things to do is to get through doorways. It had set itself a sub-goal of getting really good at getting through doorways by practicing that, which was not something that the programmers had originally envisioned.

I have written about and other people have written about how challenging this can be for law and policy, when you have these actors in the world where no person intended what they did and no person even foresaw what they did. I think of that as being really interesting and really important. I imagine those kinds of themes will come up over the course of today.

Peter couldn't join us, but I have to say that Gail Cohen and Anne-Marie Mazza have, in consultation with our program committee, invited some incredibly interesting people to discuss these repercussions with you. Topics include, of course, the economic impacts, the legal and ethical issues, the security implications, and even the inevitable international dimension of robotics and artificial intelligence. Our speakers come from the White House, from DARPA, from Homeland Security, from the German Embassy, from our flagship universities and companies, all places at the cutting edge of robotics and artificial intelligence. I truly cannot wait to hear what they each have to say.

Eventually, I am going to introduce our morning keynote, Jason Furman. Before I do that, I do want to pause to thank the program committee and to thank Anne-Marie and Gail for putting this all together. The program committee consists of myself, Bill Colglazier, who can't be here today, unfortunately - was called away - but Missy Cummings, Hank Greely, Simon Johnson, and Milind Tambe, all of whom are deeply expert in their diverse set of fields.

I also want to draw your attention and thank the Office of the Director of National Intelligence for sponsoring this. I was just recently at a DARPA event on Friday. I am always so indebted to that community for being such thought leaders and thinking so far ahead on these issues, which can be very wonderful.

I see Simon Johnson here, from MIT. Could you introduce yourself as a member of the program committee?

DR. JOHNSON: My name is Simon Johnson. I am in the Sloan School of Management at MIT. I work on economic development and macroeconomic policy, including jobs. I was previously the chief economist at the International Monetary Fund.

We have a new course that we are launching at MIT, actually starting in February, on the commercialization and policy implications of robotics and artificial intelligence. I understand there is a hype cycle in everything. The hype cycle is pretty intense. We have way more students than we can handle right now. I think that is a nice problem to have. We are hoping to feature the work of many people in this room in our discussions.

By our survey, there are 33 separate groups at MIT working on some version of AI and robotics. We are trying to integrate that. We are trying to create materials that can be used. We would love to share those materials with other people around the country and around the world, in terms of thinking about all of these issues, including the fascinating issues that Ryan already sketched out for us.

MR. CALO: I mentioned in passing the White House artificial intelligence series, one of which was called AI Now. It took place in New York City.

What the White House decided to do is to bring together experts from all over the place. It was not just the Office of Science and Technology Policy, but multiple units within the White House and to really bring together the foremost experts - many of the foremost experts on these issues in the United States. Throughout that process, I had the great pleasure of hearing from Jason Furman at one of these events. He had an extremely wise and interesting discussion about the impacts of robotics on the U.S. and international economy. I am really delighted that he is going to be able to serve as our morning keynote here.

Jason Furman was confirmed by the Senate on August 1st, 2013, as the 28th chairman of the Council of Economic Advisors. In this role, of course, he serves as President Obama's chief economist and a member of the cabinet. He served with the president since the beginning of the administration and previously held the position of Principal Deputy Director of the National Economic Council and Assistant to the President.

Like I said, in addition to having this big picture role and big picture understanding of the economy, Jason has thought quite specifically about the impacts of robotics and artificial intelligence and has been one of the drivers behind that thoughtfulness within the administration. We are delighted to have him here today.

(Applause)

Agenda Item: Keynote Address: The Opportunities

and Challenges of AI and Robotics

Is This Time Different? The Opportunities and Challenges of Artificial Intelligence

Jason Furman Chairman, Council of Economic Advisers



The National Academies of Sciences, Engineering, and Medicine Washington, DC December 12, 2016

Presentation of Jason Furman, 12/12/2016

MR. FURMAN: Thanks so much for organizing this session. I am glad that most of you are wearing ties, too. Normally, when I speak to groups about AI, I am sort of embarrassed that I am broadcasting my ignorance with a tie. I feel like I am in good company here.

This is a topic we have, as Ryan said, spent a lot of time on at the White House. We put out a report last month on preparing for the future of AI. And then later this week, a follow-up that is focused specifically on the economic implications of AI and its implications for economic policy will be coming out. To some degree, I will be able to preview that for all of you today.



Presentation of Jason Furman, 12/12/2016

I can't speak anywhere on just about any topic without getting the question of whether the robots are going to take all of our jobs. It is just all over the media. This is one of hundreds of articles that you could have read this year about how robots are going to soon take our jobs. Of course, if you were looking at the newspaper in the year 1980, as you see on the next slide, you could see just about the same thing.



Presentation of Jason Furman, 12/12/2016

If you go to the next slide, in 1960, we were talking about the robots rise, their bidding for the big jobs both in outer space, which presumably is less of a threat to most workers, and in U.S. factories.



Presentation of Jason Furman, 12/12/2016

If you go to the next slide, as far back as 1935, the Washington Post was fretting about thinking machines replacing the thinker and robot brains outdoing man's mind in speed and accuracy of results.



Presentation of Jason Furman, 12/12/2016

Then cheating a little bit and going all the way back to 1812 on the next slide, you have the thousand-pound reward for tracking down and finding any of the Luddites, who were worried about the machines taking their jobs.



Presentation of Jason Furman, 12/12/2016

So, for centuries, people have been predicting that new technologies would replace their jobs. In a sense, for centuries, they have been right. The Luddites weren't wrong. They were skilled craftsmen and their jobs were replaced by the machines. Many of the jobs that the New York Times and the Washington Post were talking about in 1935, 1980, 1990, and in 2016, indeed, were replaced by automation, were replaced by machines. Yet, as we see in the next slide, for at least the last 115 years, the unemployment rate has generally been about the same thing.



Presentation of Jason Furman, 12/12/2016

It fluctuates around it for cyclical reasons, but broadly speaking, it doesn't have a trend.

There are four reasons why the unemployment rate hasn't had a trend. The first is that as people have better technologies, they get richer. When people are richer, they want to spend more and they want to buy more. When people want to buy more, that creates jobs throughout the economy. It may create jobs in sectors completely unrelated to the automation, as people go to restaurants more or hire tennis pros or consume services across the spectrum.

The second reason is that to the degree that machines complement people's skills, with every additional hour of work, you can get more for that hour of work. That gives you more of an inducement to work.

The third reason is that new technologies can directly create new types of jobs, for example, computer programmer is a job that didn't exist in 1901 and exists in large measure today. It opens up directly new jobs.

Then the last reason, which is related to the first three, is there are just, at least to date, are a lot of things that machines can't do, whether it is general intelligence, creativity, judgment, ability to have human interaction, a whole range of things that we need to do that machines can't do.

To a first approximation, this is true across several millennia of technological progress. There aren't an awful lot of farmers today because a tiny number of people can make food compared to it used to be just about everyone in human society was engaged in food. You see this in what economists would call the time series. You see it across countries. It is not the case that countries that are more technologically advanced have a higher unemployment rate or fewer people working. The unemployment rate is pretty much unrelated to your degree of technology.

Instead, to a first approximation, what technological advance does is shown in the next slide, which is raise the GDP per capita, which has been happening steadily across all of the advanced economies since 1950.



Presentation of Jason Furman, 12/12/2016

It also lets us - in addition to consuming more material goods, as we see in the next slide, it lets us consume more leisure as well.



Presentation of Jason Furman, 12/12/2016

You have seen a downward trend in hours across the G-7.

This is the last - I am just showing you the last 60 or 100 years. The same, as I said, has been true for millennia. The question a lot of people ask, though, is is this time different? Is there something about artificial intelligence that threatens to make it different going forward than what we have seen now.

I am skeptical about that for two reasons. One is, as we saw with those headlines, people always think it is different and yet, those four forces - people being richer and wanting to spend more, technology creating new jobs, technology being unable to replace all of our jobs, and technology helping us earn more and giving us more of an inducement to work - those four forces have always operated in the past. I think there is an enormous burden of proof on people who think they won't operate in the future.

The second reason I am skeptical is what you see in this next chart, which is this disconnect that you go to technological circles and you hear about how amazing technological progress is and then, if we advance the chart, then you go to a macro conference and you are discussing why it is that productivity growth has slowed so much. This shows productivity growth from 1995 to 2005 and 2005 to 2015.



Presentation of Jason Furman, 12/12/2016

I am just showing it for the G-7. If I showed you the rest of the advanced economies, all but one of them would follow the same pattern of slowing, on average, going from two percent productivity growth from 1995 to 2005 to one percent in the last decade.

To me, this says we have seen almost the opposite in the last decade of what the AI worry was. We have seen very successful job growth. We have seen much less successful GDP growth, which is to say output per hour has not grown as much as it has in the past. Output per hour of human work is precisely what you would expect to be rising very dramatically as robots and AI were taking our jobs.

Now, there is a lot of causes of this. We could talk about that another time. One question often comes up is this just that we don't know how to measure all of the technological advances? Are these numbers missing our productivity growth? The way we could be missing productivity growth is if, for example, Wikipedia gives you something that you are not spending money on. Maybe it doesn't show up in GDP and then it doesn't show up in productivity growth. Or we might be getting quality adjustments wrong. We think the price of our smartphone went up, but really the quality got so much better - it replaced GPS, camera, eight other things - so we should actually think of the price as having gone down and the quantity having gone up.

I have no doubt that if you look at those orange or whatever color bars those are that there has been mismeasurement in the last decade and that in the United States, true productivity growth has been higher than the one percent that you see there. I think there is also no doubt that there was mismeasurement from 1995 to 2005 and that blue bar should be even higher than what is shown there as well. Economists who have tried to carefully

quantify this would estimate that if anything, the blue bar - you should increase it even more than you increase the orange bar. The gap between them, which represents the slowdown, might be even larger than what is shown there.

Just do a thought experiment, thinking about your own lives. In 1995 or slip it a year or two earlier, make it 1993, we didn't have a graphical interface for the world wide web, which is to say virtually no one used the internet. We didn't have cell phones. If you were at a restaurant waiting for a friend and they were 15 minutes late, you would go find a payphone, call your home answering machine, key through a bunch of buttons, and wonder if they left a message explaining to you the change of plans. Once you had finished that, you would then make another call from the payphone to their answering machine in the hope that maybe they would call in and find out what was going on with your plan.

By 2005, the end of the period covered by that blue bar, we not only had cell phones, smartphones were becoming increasingly common. They weren't super smart, but they could do things like email. We all used Google. We all used Amazon. We all used eBay. Our lives were really changed by the internet in a lot of ways that weren't measured.

Then you do 2005 to 2015, you know, or cheat and make it 2010 to 2015, like personally, I really love Waze. There are a bunch of other things that my mobile phone does, but it is all probably less important relative to what you saw in that decade from 1995 to 2005. In many ways, if you are adjusting for the things that we are not capturing and we are not measuring, you want to do it everywhere. By the way, it is not just technology. It is the introduction of the automobile and a range of things in the past were understated in the economic statistics. I don't think that mismeasurement is the problem here.

The reason that these data don't track what most of us have in our heads or at least in our heads if we are staring at our smartphone is that technology is still a relatively small part of our economy. It is affecting a lot of things, but a lot of what we do is healthcare and building houses and educating people and feeding people in restaurants and helping them with their tourism and a whole range of things like that that just aren't changing at anywhere the same pace. You have a small fraction of the economy growing really rapidly, a lot of it growing much less rapidly.

To give you an example, one of the technological marvels that is showing up in the orange bar and maybe is

being mismeasured is things like ridesharing services. That has changed the way a lot of people get their transportation. Just to quantify that, air travel we spend 15 times more on than we spend on all taxi, limousine, ridesharing ground transportation combined. Air travel, you could sort of debate - I think it depends on which part of the plane you sit on, whether we have seen quality improvements in the last decade or it has gotten worse. Things like planes, on average, used to have half the seats filled. Now, on average, they have something like 90 percent of the seats filled. You used to, in effect, be buying an empty seat next to you. Now, you are buying some person next to you. That is an unmeasured way in which quality has actually gotten worse in air travel. I think if you added all of that up, it would be quantitatively potentially larger than what has gone on with things like Uber, just because the air travel sector, itself, is so much larger.

Now, we could discuss whether our future is going to look like those orange bars or pop back towards the blue bars. I think there is a lot of reason to be excited about the technology and innovation we have seen and think it will take some time to come into force. I think part of why those orange bars have been like that is the fallout

from the global financial crisis, which led to less business investment. When you have less business investment, you have less productivity growth. That fallout won't go on indefinitely. We are getting increasingly beyond it. So, I think to some degree, it is idiosyncratic and won't continue. To some degree, it doesn't reflect the promise going forward. To some degree, I think it should be a little bit sobering.

Now, this isn't the fault of AI and robotics. That has been trying to help, as we see in the next slide.



Presentation of Jason Furman, 12/12/2016

This is industrial robots, which are obviously, of course, one way in which AI can manifest itself and replace one type of task. We have seen a big increase in shipments. This is just something we can easily measure. Economists who have tried to quantify it have found that across a range of countries from 1993 to 2007, the increased use of robots in production added 0.4 percentage points to GDP growth or to productivity growth. This has been part of the solution, it just hasn't been big enough to overcome what we are seeing in all of the other parts of the economy.

I would summarize this first part of my discussion as saying to a first approximation, the problem we have isn't too much AI replacing all of our jobs. It is not enough AI. If we had more AI, we would have more productivity growth. If we had more productivity growth, we would be able to choose some combination of higher spending on consumer goods or increased consumption of leisure, so people working fewer hours. Those would be a wonderful set of choices to have. A lot of our worry as policymakers should be what can we do in order to have more AI and see more of this thing that some are worried about. I think to a first, second, and third approximation, our reaction to getting more AI should be to celebrate it and

be excited about what expanded productivity growth should do to us - can do for us.

Once we are done with our first, second, and third reaction, which would be celebrating, we can get to our fourth reaction, which is to be concerned about some of the side effects. I want to show you how I think about some of those side effects, stressing that they are not technologically determined. They are very much a function of the institutions you have in your country and the policy choices that you make. They are not something that is directly caused by the technology, itself.

The backdrop for this is in the 19th century, we had a lot of technological progress. A lot of what it did was take skilled artisans and replace their jobs and instead, take people who are unskilled and could work on a factory assembly line and create not just a job for them, but a job in which they could produce a lot more per hour than they did before. This unskilled bias technological change - technological change that helped people without skills helped reduce inequality.

Over the last several decades, we have seen the opposite. The technology has replaced routine tasks, repetitive tasks, things that are generally done by people with less skills. That has caused a reduced demand for

them and thus, higher inequality. At the Council of Economic Advisors, we tried to assess what AI might do for this going forward. For that purpose, we used a paper by Frey and Osbourne at Oxford, who estimated that 47 percent of the jobs that we have in our economy could be replaced by automation. Now, I don't want to have to debate their numbers. The OECD looked at it and put the number at nine percent. I want to use this more in a relative sense to look at the impact at different parts of the wage distribution. Even if you don't believe the 47 percent, it is a way of calibrating where the robots will be competing with us. The next slide shows quite dramatically.



Presentation of Jason Furman, 12/12/2016

So, the Council of Economic Advisors, what we did was we took the Frey and Osbourne list of occupations and we divided them up. People that made less than \$20 an hour, 83 percent were subject to automation. \$20 to \$40 an hour, it is 31 percent. More than \$40 an hour, it was only four percent. People have the examples of the lawyers and the radiologists in that four percent, but there are still many, many more people who are paid less than \$20 an hour who have occupations that when Frey and Osbourne looked at it, they thought were subject to automation.

Now, again, the right way to think about that isn't that 83 percent of the jobs making less than \$20 an hour will disappear. Part of why they won't disappear is those robots will compete with those people. Those people's wages will go down. That will make them cost competitive relative to the robots. At more than \$40 an hour, not only are you not competing with the robots, but much more likely your skill is going to be complemented and expanded by them. It is going to enable you to do more than you are doing today. As a result, you will be paid more. Certainly, all of the owners of the robots would benefit from all of this as well. So, the same forces that we have seen increasing inequality over the last decade are likely to continue and potentially continue in a more
extreme way.

The next chart I have is based on a different paper, which has a different average percentage of people who have jobs with highly automatable skills.



Presentation of Jason Furman, 12/12/2016

It tells exactly the same story, 44 percent of jobs held by those with less than high school are highly automatable, zero percent for those of us with a graduate degree.

The reason, though, that I don't think this uniquely gives us the answer to technology causes inequality is what you see in the next slide, which is the amount of inequality - this is the share of income earned by the top one percent - varies a lot across countries, from 18 percent in the United States to more like eight percent in France, Japan, and - I don't know if that is Italy or the United Kingdom down there. I think it is Italy.



Presentation of Jason Furman, 12/12/2016

That isn't a measure of your technology. That is a measure of a range of institutional choices.

The story of the last couple decades, I called it skill-biased technological change - I mean I use the term economists use - skill-biased technological change, but that only tells us half the story. That said the demand increased for workers with a college degree. At the same time, in the United States, the supply of those workers slowed down. We used to add about .4 years to the total number of years of schooling every five years. Now, we are adding about .15. We are adding schooling at about a third the rate that we used to. Part of what has gone on in the United States is the demand for skills has gone up, but the supply of skills has gone down or at least it hasn't gone up as fast. It is that combination that has resulted in increased premium.

Also things like large-scale de-unionization in the United States. Less for the top one percent, but if you are looking towards the bottom of the distribution, the fact that our minimum wage is the lowest of other countries - there is a whole range of institutional and policy choices, whether they are how you divide the income or what skills you give people, that have made very different trajectories of inequality across countries and very different levels today. The first worry is inequality. This will put pressure on it, but it need not be inevitable.

The second is, to give a slight qualification to the argument that I made at the beginning - at the beginning, I showed you the unemployment rate for 115 years

had basically not changed. If you look, though, at the percentage of men participating in the workforce - that is men who either have a job or are actively looking at a job, that actually has fallen quite continuously. In the 1950's, 98 percent of men between the age of 25 and 54 were either in a job or actively looking for a job. Now, it is 88 percent of men. That is a ten percentage point decrease. To give you a sense, ten percentage points often during a recession the unemployment rate goes from five to seven. This is 5x a recession happening over a longer period of time.

Women followed a different pattern. They entered the workforce in large numbers up through about 2000. Since 2000, the trajectory for women has been similar to the trajectory for men as well for a lot of the same reasons.

This is not something that I think, again, we can simply understand as automation took someone's job away. We have huge churn in the economy. Every quarter, the economy adds about 500,000 jobs. Every quarter, what actually happened was the economy added six million jobs, but it also lost five and a half million jobs. As a result, employment went up 500,000. There is huge amounts of churn in the economy. Some of that is automation, but a lot of that is just businesses starting and failing.

In the United States, we have been less good at getting people who are subject to that churn into a new If you look at the next slide, it tells that story. job.





Presentation of Jason Furman, 12/12/2016

surce: Organization for Economic Co-operation and Development

You look at the U.S. labor force participation rate or the red bar. This is the percentage of men between 25 and 54 in the workforce. We are a tiny bit above Italy, above Israel, and lower than every other one of the advanced economies. This, again, is not a measure of how much AI you have in your country. If you compare Switzerland and Italy, the difference isn't that in Italy, there are lots of machines doing the jobs and in

Switzerland, they are all making the cuckoo clocks by hand. This is a reflection of institutional choices, about how you organized your economy and how you set up your labor market. One of them being, for example, in the United States, our what is called active labor market policies, helping people to find jobs, helping subsidize them getting into those jobs, helping train them for those jobs, we spend 0.2 percent of GDP on that. That is lower than every one of those countries that you see on this slide except Mexico and Chile. It is about one-third the average in the OECD.

Part of the story of the last 50 years has been this shock that has hit all of these countries in terms of technology. That shock is mediated by your institutions and by your policy choices. As a result, it has resulted in a larger increase in inequality which rose to a higher level in the United States and also a larger decline in labor force participation which fell to a lower level in the United States than most other countries.

This all brings me to policy. Start with something that I am much more skeptical makes sense and that is universal basic income (UBI). It doesn't make sense for three reasons. It is a little bit hard to debate with universal basic income because it keeps changing how

it defines itself. I am assuming it means universal, which is everyone gets it. I assume it means basic, which is you all get a certain amount and that it is income, that it is in cash.

The first issue with it is a separate debate, which has nothing to do with automation and AI, which is just how well designed or not is the safety net today. I think the safety net today is in need of improvement or could benefit from improvement, but that it also does quite a good job lifting a lot of people out of poverty. If anything, there is a recent range of research that has found that programs ranging from SNAP, Medicaid, to housing vouchers, actually result in very large long-run gains for the children that get those programs in terms of education, health outcomes, and earnings. That is a separate debate we could have and sort of a public finance debate.

When it comes to AI, I think the problem with UBI is it is based on the premise that you are basically giving up on people working. Instead, you are giving people money, in effect, not to work when I see nothing from the last millennia of history from the comparison across countries to say that people can't work. There is an enormous amount of evidence that when people don't work that they have lower self-esteem, that they are more likely to use opioids, suicide, or die of natural causes. So I think the focus of our policy should be to be how do we navigate this transition in a way that helps people continue to get jobs and be in jobs, not, in effect, give up on that.

UBI also has some arithmetic challenges, which is if you tried to make a substantial amount, say \$10,000 for each person and you add it up, you would discover we have a program that is several times larger than what is currently our largest program, Social Security. If I thought we could raise taxes in the United States - could or should raise taxes in the United States by five or ten percent of GDP, then I think this could be a discussion that one would I don't think we will ever do that, which means I have. don't think we even need to debate whether we should do that. That means the only way to have this UBI would either be at the expense of other things that we really need like infrastructure, basic research, or at the expense of the existing social safety net. I think when you try to actually write down a particular model, it turns out that it tends not to add up and work in the way you want.

What do we want to do? The next slide gives you a brief version of that.

<section-header><section-header><section-header><section-header><list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item>

Presentation of Jason Furman, 12/12/2016

The first thing, as I said, is we need more AI. We should be excited about more AI. We are doing a lot of research. Businesses are investing in basic research - R&D as a share of GDP is at the highest that it has ever been. We have record business investment. At the same time, federal investment has been trending down since the 1960s. The federal government still does most of the basic research where there is less of an incentive for a company to do it and more spillovers and positive benefits to the rest of the economy. Increasing that is particularly important. Some of it need not be expensive. It is also how you use the money, things like the role that DARPA's Grand Challenges played in stimulating investment in autonomous vehicles.

We need a workforce to develop AI. AI doesn't program itself. There are a lot of things in workforce development. Competition is really important to innovation. There is a lot of aspects of technology that lend themselves to network effects, which can reduce competition and ultimately, create concerns around innovation. For people to trust AI and be willing to use it, we also need to do more for privacy and cybersecurity.

When I said number one, two, and three is we want more AI, number four, we are concerned about side effects, a lot of the side effects aren't novel, new things opened up by this technology. They are the same side effects we have been worried about and seen for the last many decades in technology. A lot of it is the things that we should be doing already, just with additional urgency because of this - moving people on that education gradient towards where they can't be replaced by the robot, helping active labor market policies, empowering workers and job seekers in ways that would help reduce inequality. There are ways to strengthen the social safety net programs like wage insurance that would replace a share of your lost earnings if your job was replaced, and then tax and economic

policies more broadly.

In conclusion, I would say I think this is a really exciting area. It probably will contribute more to our economy in the future than it has in the past. I think it will continue us along a continuum, one that offers potentially a lot of benefits, but it depends on the policies that we take to seize and make sure we are enjoying those benefits. Thank you.

(Applause)

DR. JOHNSON: My name is Simon Johnson from MIT. Thanks very much, Jason, for coming today. Very helpful and enlightening remarks. Great way to start the day.

I completely agree with your point about unemployment not having a trend. Of course, there is another narrative or another way to construct the narrative, which is about job polarization, in which automation and information technology would be very central and over a long period of time, we have seen not unemployment, but we see many people in middle income/middle skill/middle class jobs becoming displaced and moving down to lower wage, lower income jobs. We did try to have David Autor and Daron Acemoglu to join us today, but they weren't available. That would have been one of their themes, I'm sure. I wonder if AI is continuing that or reinforcing that job polarization, which we will talk about, which seems likely.

Perhaps we should be considering more dramatic responses to it. I don't disagree with any of the agenda there, but I wonder if, given our starting point, for example, something more dramatic on education - I would be interested in your thoughts on that - something more dramatic on taxation. Why do people earning below \$50,000 a year pay any tax or need to pay any tax, including Social Security contributions in the world going forward? That would be a fairly radical departure from the way we thought about social insurance and how you contribute to it, but I wonder what your thoughts are on those potential policy responses.

MR. FURMAN: Sure. I think in some sense more dramatic is better. I have spent like eight years dealing with one's inability to always get things dramatically done. If you invite me back six months from now, I am sure I will be helping you all expand the Overton window, but I think those are exactly the directions. I don't think we have any disagreement at all that that is exactly the right direction.

You look at the gradient in terms of where you are likely to be replaced by AI versus where you are likely

to complement it and the best insurance you can buy is more education. Does that require more federal dollars or should we consider some form of an expanded and simplified income-based repayment for student loans? A lot more people are borrowing money for school. Could you go all the way to the Australian model and make it even easier and more automatic for people so that, in effect, you are financing college entirely with loans. You are forgiving loans for people for whom it doesn't pay off. You are having the people for whom it does pay off pay that back. You are not on net, increasing the subsidies to the sector. I think some big ideas like that that need not cost money.

In terms of the tax system, I think absolutely, you want to make it a lot more progressive. Right now, if you make less than \$50,000 a year and you have children, chances are you are not paying taxes because of the EITC and the child credit. If you are a childless worker, we actually tax you deeper into poverty. If you are already below the poverty line, we tax you and make you even deeper into the poverty line, as opposed to with children, we are much more likely to use the tax system to get you above the poverty line. There are some things which are compositional. I think the first set of dollars I would use are on those childless people. So you had a principle

that if you are in poverty, we are not taxing you at a very minimum.

Then we can also afford to - if you look at our budget, we don't raise rates on high income households. We got about a percent of GDP a year in additional revenue just from closing loopholes and a number of things, many of which would actually improve the efficiency of the tax code by reducing, for example, the incentive of the top one percent to over-invest in housing that they have today through the tax code. I think there is a lot that could be done to the progressivity and then you can debate rates after that.

DR. NITZE: Bill Nitze with George Mason. Can you talk a little bit about geography and culture? Just reading today about the difficulty the Republicans are going to have in replacing Obamacare. One thing came out and that is one of the problems is the irrational resistance to certain types of transfers in assistance in red states, which will, ironically, greatly increase the impact of some of the things that the repealers would like to do in those states. The question that raises is - I can see in a state like New York or California where many of the proposed policy developments that you would wish would be received well. I suspect that it is a harder lift in

Appalachia for cultural reasons. I also see the possibility of a great discrepancy in impact, both from AI, itself, and from policy measures to deal with the issue of income inequality. I just wonder from your eight years, if you could reflect on these geographical and cultural issues.

MR. FURMAN: You are talking about an increased set of rewards for knowledge. Knowledge tends to do a lot better in proximity to other knowledge. If it is natural resources, you need to be near the natural resources. Τf it is factories, you need a cheap place to build a factory. If it is knowledge, you need to be around other knowledgeable people. Nothing in the technological developments we have seen so far, witness all of you here in Washington, D.C., in this room, have replaced the role that face-to-face interaction has. This guy, Richard Baldwin, has an interesting book where he thinks the next wave of globalization might be telepresence and other things. Videoconferencing will be good enough, you will look like you are sitting right there. You won't actually need to make your way all the way here. We haven't seen that to date.

One thing that has done is it has concentrated a lot of the benefits in a set of areas. A lot of those

areas have then decided they don't really want more people moving to those areas so they create land use restrictions that make it more expensive to build housing. Some of the process that both is important for productivity growth, of letting people agglomerate where they are most successful, but also for people doing better, which has always, in our history, involved mass migration, has been slowed a little bit as people put up gates in a way that they weren't able to do 50 years ago. There weren't land use restrictions that prevented people from moving to the Sun Belt and jobs or earlier than that, moving north to factories. I think that interaction of the way we handle land and mobility is something that is a little bit of a nagging worry for me going forward.

Certainly, there are also, yes, big cultural and other differences. People have seen the election results broken down by GDP and how skewed they were in this last election relative to previous ones. I think that is skewing along educational lines, those being correlated with place and all of this is very much related.

I'm fine for time, but I don't want to take away from your next panel so you should just throw me out.

MS. MULLIGAN: Deidre Mulligan, Berkley. I was wondering if you could talk a little bit more about privacy

and cybersecurity. They were up there. The state of research in both of those areas, I think, is far more advanced, beyond what we actually see deployed in the marketplace. I think we are at a pretty important inflection point as we are embedding technology and things that are moving about. The voting system issue I think has everybody quite focused on earlier failures, which I think have been about both market and regulatory structures. I am wondering if you could talk a little bit about those issues.

MR. FURMAN: Sure. There are others who are far more expert in them. To the degree I have given thought to it, it is what are the role of the government in those I think most of the important things are what the areas. private sector will do in those areas, the private sector has a real incentive to do something in terms of the reputational damage that it does, but it doesn't have a sufficient incentive, of course, because of spillovers. Ιf one company gets hacked, that to some degree, reduces the reputation of everyone in that sector. There is also important public issues of you build a database - that relate to the competition issue around this as well. You could get more competition and maybe more innovation if more information is shared or available, but that can come

at the expense of privacy. I think there are some interesting tradeoffs that places like the FTC are exploring at the nexus between those.

A lot of our approach in this administration to things like privacy has focused more on a light touch and a multistakeholder approach and getting people together to try to work things out. A lot of that is very slow. It is like apps for children under 13 and you spend two years in a room talking about that and you think you maybe figured that out and then move on to the next topic, as opposed to something that is more global.

For cybersecurity, one issue is just we are underinvesting in it, both the research into it and deploying it at the federal level. We have called for a large increase in federal spending on it. The second is there is where you, in particular, have this disconnect between the incentives that a company faces and the consequences of their not doing it. Certainly, in a lot of important sectors of our economy, having, in a sense, different requirements and regulating that and the way we regulate many other aspects of security and safety makes a decent amount of sense. Again, regulating more on outcomes than the exact way to do it because we certainly don't know enough to tell every company the exact way to do it. There

are certain goals, though, that we would want everyone to achieve, so being goal-oriented in that regulation.

Anyway, thank you so much.

MR. CALO: Thank you very much, Jason.

(Applause)

MR. CALO: We are going to welcome up our panel and your moderator, Simon Johnson. Welcome up the first panel here, today. Thank you.

Agenda Item: Panel 1: Economic Impacts: Implications and opportunities for growth, economic mobility, and inequality

DR. JOHNSON: I guess we are making an immediate transition to our first panel. The topic is Economic Impacts: Implications and Opportunities for Growth, Economic Mobility, and Inequality. My name is Simon Johnson from MIT. We have - just going down in this order and this will be the order of our speakers - Caroline Atkinson from Google, Guru Banavar from IBM, Martin Ford, author and futurist, Milind Tambe from University of Southern California.

Guidelines - I asked everyone to aim for 10 to 12 minutes. I am not cutting anyone off until 15 minutes. I will try really hard to cut you off at 15 just so we can get some interaction with the audience. Right, Caroline,



Presentation of Caroline Atkinson, 12/12/2016

MS. ATKINSON: Thanks very much. Let me just introduce myself briefly. I work at Google and I am head of global policy there. Until the end of last year, I was President Obama's Deputy National Security Advisor for International Economics. I have been a policymaker and worked, obviously, with Jason before that.

Just to let you know and you may have seen, we, at Google, think that artificial intelligence, AI, and machine learning can be a very significant impact on technology and innovation. Somebody was asking me last night, well, are things really speeding up? That goes to the point that Jason was making at the beginning. We haven't got enough AI. Innovation and productivity growth have been slowing. So how optimistic are we? We are optimistic. My boss, the CEO of Google, Sundar Pichai,

believes that there is enormous potential.



Presentation of Caroline Atkinson, 12/12/2016

He said, we moved from a desktop world to a mobile world. Next is an AI-first world.

We are actually using AI across all of our products. I will just give you one example here of where we have seen an enormous improvement in productivity in the area of translation.



Presentation of Caroline Atkinson, 12/12/2016

You can see that in a whole lot of languages or in all languages, human translation is much better than a machine translation. That is particularly true for difficult, if you like, or very different languages, English to Chinese, Chinese to English. There was a big improvement between - the blue line shows you phrase-based machine translations, where machines would take a different piece of the word and the phrase and translate that and then agglomerate it to what we are working on now, which is neural machine translation, where the machine figures out the context. This is particularly important in a language such as Chinese moving to English because the context matters very much, as those of you who know Chinese, to the meaning of the words that are being translated. This has

led to a reduction in the errors made by machines of between 55 and 85 percent. As you can see on the chart, machines have got a lot closer to humans.

We also, as I said, see machine learning and AI as valuable across products.



Presentation of Caroline Atkinson, 12/12/2016

It is not a separate issue. This shows you since 2012, when the breakthroughs began on this deep learning neural networks, an explosion in the use in Google of machine learning across our products. We have projects in all of these areas, whether it is Gmail, Maps, Photos, et cetera, that are using machine learning now. We expect that that will lead to a big increase in innovation and efficiency of all of our usual Google products.

We also have been working a lot on new products.

This shows a graph of another thing that has attracted a fair amount of attention.



Presentation of Caroline Atkinson, 12/12/2016

When we have been using machine learning and others have been using machine learning to drive progress, for example, in health analytics. This is a way of showing that diseases, such as diabetic retinopathy, which machines are becoming better at diagnosing than radiologists, human radiologists. That can be an example. That is an example of the enormous improvement that you can get in many other fields than just how your Gmail or your Maps work.

We also know that there is a lot of uncertainty. As Jason outlined just now very eloquently, there is a danger that the next wave of technological innovation powered with AI will drive inequality, which we already know is a big problem here and elsewhere. He showed a very nice chart about how the returns to education are going to get that much larger, essentially, with jobs that - with people earning less than \$20 an hour were much more likely to be affected by the introduction of AI and future learning processes from machines than the better paid jobs.



Presentation of Caroline Atkinson, 12/12/2016

This chart shows the probability of computerization of different kinds of jobs. It is not very clear on this, but on the right-hand side, the bars there this is the 47 percent that he referred to from the Frey and Osbourne paper. The 47 percent are mostly low-skilled employment. On the left-hand side, if professions and skills that will be less likely - somewhat less likely to be affected are more of sort of high-skilled, but still non-professional. In the middle, the impact on employment of AI in - on jobs that involve high skills and high education is that much less. This is a very important need, as Jason laid out, for policy choices.

There is also an important way of thinking that others in this field have laid out. That it is not just about jobs. It is about portions of jobs. We know that the job of a clerk used to involve a lot of writing 100 years ago and now it involves very different things. You have a portion of - we all probably, to a greater or lesser extent, have portions of our jobs that can be automated easily. It is important to make sure that we remember that although portions of our jobs are now more easily automatable, the whole job is not necessarily automatable. It can be just a task. Actually, one nice example, which is not this one, but a lot of people would have thought when banks introduced automatic tele-machines that there would be a big decrease in the demand for bank tellers. In fact, there are more bank tellers now. They are doing different things from just handing out the notes, which you can now get from automatic machines.



Presentation of Caroline Atkinson, 12/12/2016

This slide is a reference to our open source AI product, Tensor Flow, which we have decided to make open source, as we have done with Android and other innovations at Google. Already, some businesses are starting to use Tensor Flow as a way to incorporate AI in their production. There is a nice example of this guy, who is a Japanese cucumber farmer.



Presentation of Caroline Atkinson, 12/12/2016

They incorporated into their cucumber farm Tensor Flow, which helps them to sort cucumbers automatically by size, shape, and other attributes. There is also an application in a lettuce farm where a lettuce bot has been developed that is able to distinguish between lettuces and weeds and just use the weed killer on the weeds and not on all of the lettuces. There are many ways that there can be practical implications of AI in existing businesses.

It is also possible - I mentioned the health impact, but we have also worked at Google on the big issue of energy use and climate change. We have - this shows this is a picture of inside of one of our data centers.



Presentation of Caroline Atkinson, 12/12/2016

As you, I'm sure, all know, data centers for companies such as ours are now amongst the biggest energy consumers. We applied machine learning to figure out how to cool our data centers more efficiently. That has enabled us to reduce energy consumption by up to 40 percent. Also, we just announced last week that we will reach the goal in 2017 of 100 percent renewable energy sourcing our data centers. That is helped partly by big projects that we have had of supporting wind farms and so on, but also by the more efficient usage of energy. Obviously, that has got major impacts elsewhere.

Going back to the question of how can we mitigate the risks - I mean the bottom line is that we are very excited about AI. We think it can be used across a load of our own existing products, but that also it can get much more broadly applied.

One of my colleagues, who you may know, who runs Deep Mind, which was taken over by Google and was the organization that used artificial intelligence and machine learning to figure out how to play the game of Go, which is many times more sophisticated than chess, and Alphago, that Demis has developed. There was a championship between Alphago and the Korean champion of Go in March this year, where the computer or the Alphago program won four out of five games. That would not be possible by advanced programming. It was the machine learning of the program or of the software. Anyway, Demis gave a demonstration or gave a talk recently that I was at where he was saying, well, what are our goals? Our goals are, number one, solve intelligence, number two, solve everything else.

Despite that rather exaggerated claim, he is actually a very nice and humble guy. He did say to me, your stuff, you know, politics, economics, policy, is the hardest thing. We will be able to deal with harder science - is the most difficult things. We will be able to deal with harder science first. That does mean that there is a lot, as Jason pointed out, that policymakers and institution builders need to do to make sure that the

impacts on society are ones that are good and what people want to see.

Jason, again, outlined and Simon asked him about the - where should policymakers focus. Education is obviously a huge area.



Presentation of Caroline Atkinson, 12/12/2016 I think there is a lot more that can be done

about improving access to digital skills.

The National Academies of SCIENCES • ENGINEERING • MEDICINE



Presentation of Caroline Atkinson, 12/12/2016

There is also a hope that events like this and the work that the administration has been doing most recently can help us all to figure out better policy solutions to explain why they matter and how AI can be used - can be harnessed and machine learning can be harnessed to help improve lives more generally. Thank you very much.

DR. JOHNSON: Thank you, Caroline.

(Applause)

DR. JOHNSON: In 13 minutes, brilliant work. Guru is next from IBM.



Presentation of Guru Banavar, 12/12/2016

DR. BANAVAR: Thank you. Good morning. My name is Guru Banavar. I am the Chief Science Officer - so I actually have one foot in the science and one foot in the business of AI. We think about these technologies as cognitive computing systems that augmented our intelligence, which is a broader thought than artificial intelligence alone. From an economic perspective, we look at this point in time as a natural evolution of how we have augmented our physical abilities with tools of different kinds throughout our history.

I think we are in an era where we do need cognitive assistance, where we are overwhelmed as knowledge workers in the economy with the amount of not just the data, but the knowledge that is being generated, which, by the way, is also growing exponentially. For example, when you go to a typical doctor, you will probably see that they are out of date by five to ten years. How do you get that doctor up to speed with all of the latest clinical trial research? That is the topic of my discussion because we think cognitive computing is going to be that assistance – for your cognitive abilities as you make decisions in your daily life. So, this is my video, which I promise you has relevance to the discussion today.

(30 second video of Ken Jennings talking to

Watson about what it has learnt since that famous TV quiz show - available at:

https://www.youtube.com/watch?v=Keq0XWpBEoQ)

DR. BANAVAR: While this is a tongue in cheek type of a video, I wanted to highlight two points here. One is the examples that Ken Jennings used about retirement planning and cancer treatment. Those are the kinds of jobs and occupations that are going to be significantly augmented. In fact, those workers will become much more effective because of the use of cognitive computing. The second thing is about the limitations of these cognitive computing machines, where they talk about sarcasm in the video. Really, there are a whole bunch of skills and capabilities that are absolutely and centrally human which are needed in many of these jobs. Some of them have to do with empathy and understanding the social context of people. Others have to do with common sense. Think about management. Think about creativity and those kinds of things, which are extremely relevant and important to the discussion we are going to have today.

This next slide shows the amount of transformation that is going on in every single industry.



Presentation of Guru Banavar, 12/12/2016

IBM works in every industry from health all the way to manufacturing and education, even government as an industry of industries. There is a huge amount of raw data, and there is also a tremendous amount of curation of that data that is required in order to create usable knowledge – as we extract features or link the various concepts within the raw data and then find a contextual way of representing it. That process is absolutely essential for making the right kinds of decisions. In fact, if you look at something like healthcare, most of the data that we have available today is not actually being used for making more effective decisions. Turns out only about 25 percent of the data that somebody might need for making the best decision from a healthcare perspective is actually digitized. 75 percent of the data is not digitized because it has to do with our social context, our psychological makeup - your stress levels, your posture, and things like this, which are not available in digital form today. Even if some of it is available, it is not used by the healthcare system, which is why we are paying the price for not knowing what is going on with a patient, with a student, with our environment.



Presentation of Guru Banavar, 12/12/2016
I think more AI is needed in order to make this work.

As I said before, cognitive computing is creating these partnerships between people and machines to make these decisions in a much more effective way than either one of them can do alone.



Presentation of Guru Banavar, 12/12/2016

I am going to show you a few examples.



Presentation of Guru Banavar, 12/12/2016

Some examples come from cancer research and cancer practice. There was a 60 Minutes segment just about a month or so ago that I would highly encourage this audience to look at where we have Watson as a member of a tumor board, a cancer tumor board at the University of North Carolina Chapel Hill. We tested over 1,000 cases where Watson came up with the same set of treatments, after having gone through the literature and the other drugs that are available in the specific cases, more than 98 percent of the time, Watson came up with a good and actually a similar treatment regimen as the doctors on the tumor board. In 30 percent of the cases, Watson came up with additional recommendations that the tumor board had not considered. That was actually eye-opening for even the M.Ds. on the tumor board. There is going to be a publication on this study shortly.

Medtronic has recently built an application for managing type 1 diabetes over time through your smartphone, getting sensor data to track a patient's glucose level and helping you manage potential diabetic events and avert serious consequences. We worked with Sesame Street for education-related applications where you have children working with Watson question and answering to get kids learning in a much better way. We even have tourism and sports applications as well.



Presentation of Guru Banavar, 12/12/2016

The latest one here is an application that built a trailer for a movie semi-automatically. This is not 100 percent automatic, but the machine actually develops a bunch of potential exciting events that are sort of strung together in a set of storylines that a person, a designer could then pick pieces for a very interesting trailer that you can, in fact, see on YouTube right now and decide whether you think it is a good trailer for this movie called "Morgan".



Presentation of Guru Banavar, 12/12/2016

All of these examples should tell you that every single industry has applications that can be built on the Watson platform. The purpose of the Watson platform has been to lower the barrier to entry for any startup, any institution, anyone, even students, to build applications using a cloud-based API where you have basic components of AI that are already available. You can build applications on top and these components available on a cloud using data and models at the bottom. As we speak, there are literally tens of thousands of application developers who are building their applications on top of this platform.

I bring this up because I think it is very relevant to

how industries are being transformed and how jobs are going to be, in fact, impacted. I believe that jobs are going to be impacted in a positive way. I am going to make that case in the next few slides.

Our vision, of course, is that every single professional in the future will have a cognitive assistant to help them do their daily tasks.



Presentation of Guru Banavar, 12/12/2016

You can be a medical professional, you can be an education professional, a businessperson, or any job in a company like a salesperson or a marketing person and you will have a - think of a Watson on your shoulder whispering to you about what are the options available to you as you go around making your daily decisions on various things.



Presentation of Guru Banavar, 12/12/2016

That, we believe, is a collaborative mode of work that all of us have to get more comfortable with, collaborative, in terms of a person and a machine working together.

As I said before, the machine can actually come up with more options, give you the pros and cons for each of the options, maybe eliminate some bias in certain cases, and do broader analysis like population analysis. Whereas, the person can bring in the creative problem solving aspects, the common-sense aspects, the value judgment, social benefit aspect, and so forth. That combination, we believe, will create a much better outcome at the end of the day.

In fact, the approach that we are taking is to look at all of the occupations - this comes directly into the topic of this conversation today.



Presentation of Guru Banavar, 12/12/2016

At the very top of this chart are occupations that we picked up from ONET. There are thousands of them. You can think of engineers and scientists, but also finance people like accountants and medical professionals and so forth at the very top. As you think about how these professionals achieve their facility in their chosen professions, you can see on the left-hand side they go through a set of educational steps - elementary school, middle school, high school, graduate school, professional education, and so forth.

At the bottom, I have just categorized the kinds of tasks that they learn how to do. I have ingestion task, cogitation tasks, generation tasks, and so forth. This is not the taxonomy for all kinds of tasks, but I think it is representative to explain how we think about building these cognitive machines. When you are in kindergarten, you learn how to interpret pictures. When you go into graduate school, you learn how to write research papers. When you go into professional level education, you learn how to negotiate, etc.

How do we teach machines all of these tasks? We, essentially come up with a slice of tasks that are relevant to a particular cognitive assistant. So a radiologist cognitive assistant requires the ability to look for anomalies in images or understand certain elements of graduate school anatomy and certain elements of interpreting longitudinal patient data and puts it all together in order to provide the right kind of decision support.

As you look at it, you can think about some tasks as being more creatively oriented or more unstructured.



You can think of other tasks as being much more routine.

Presentation of Guru Banavar, 12/12/2016

In fact, if you look at the progression of the different kinds of working tasks over time, you can see the top two are unstructured and tasks related to new information. Whereas, the bottom three are manual or repetitive routine tasks. I suspect that this kind of an analysis is going to happen in every industry. The ones at the top are going to require much more human capabilities, which is where I think we need to focus our efforts around educational training. As Caroline already mentioned, there is going to be demand for new kinds of skills.



Presentation of Guru Banavar, 12/12/2016

For example, ATMs replacing tellers has actually created a new demand for a set of skills that over time has shown that tellers are now customer relationship managers or investment managers and so forth. Re-skilling and retraining have to be done in the context of every single one of the new jobs.

The last thing I want to mention to you here is that we have started a new mode of educating people focused on specific tasks. As opposed to qualifications for jobs.



Presentation of Guru Banavar, 12/12/2016

P-TECH is a vocational education program, which is a six-year program. Instead of your normal four-year high school, it is a six-year high school program, which has just created the first set of graduates, some of whom have been hired back into IBM and other companies around the US. Today, we have about 60 such schools around this country, actually around multiple countries. I think these kinds of education programs may be much more important as we look at skill-based discussions in the future.

I think that is the last slide I had. Thank you very much.

(Applause)

DR. JOHNSON: Thank you very much. Again, perfectly on target and on time. Martin is our next speaker.

MR. FORD: Thank you. As you may know, I am the author of the book, Rise of the Robots, which makes quite a different argument from the one that Jason made earlier. I am going to be kind of the pessimist and really argue for at least a more pessimistic outlook for jobs, which is normally my role in events like this.

I do believe that this time is different. That is kind of the central question here. I think if you look at the technology, there is a lot of evidence to suggest that we are really looking at something quite unprecedented. There are a couple of things, in particular, that I would point to.

The first is that the machines are, in a limited sense, beginning to think. They are taking on cognitive capability. It is not just about muscle power or mechanical things anymore. It is really about machines that are moving into that cognitive space. The second thing is that IT and increasingly, artificial intelligence is really becoming a true general purpose technology. What I mean by that is that it is going to scale across everything. It is going to impact really every sector of

the economy.

Many people have compared information technology to a utility. You know, you wouldn't ask the question what industries or occupations are most affected by electricity. That is kind of a silly question. It is everything. The same is going to be true of artificial intelligence. I think that it is going to have a much broader, more scalable impact across everything, across all types of work and all industries, including even industries that don't exist, certainly, as well.

That is quite different from what we have seen historically. There certainly have been major technological disruptions in the past. They tend to be more specific and on a sector by sector basis. The classic example would be what happened in agriculture. It used to be that at least half the people in the United States worked on farms. Now, the number is one to two percent of the workforce works in agriculture. Millions of jobs were lost there. Of course, that didn't result in permanent unemployment. People - there was a short-term disruption, but eventually people moved on to other areas. They moved to factories. Later on, they moved from factories to the service sector as factories automated. You have seen sort of this sector by sector impact. When those disruptions

happened, there was really the whole rest of the economy out there to absorb all of those impacted workers.

This time around, what we've got is this generalpurpose technology that is basically going to scale across everything, more or less simultaneously. Some sectors will be impacted ahead of others, certainly, but in a more or less simultaneous fashion. It is really unclear what new sector is going to be out there to arise to absorb potentially millions of workers that could be impacted. We are going to have things like self-driving vehicles. We are going to have self-driving trucks and cars. We are going to have robots that, I think, dramatically impact the fast food industry. Just those jobs right there, you are already talking about millions of jobs. It is potentially a very substantial impact.

Now, historically, as I have said, what we have seen is that workers have adjusted. They have moved from sector to sector. They have really, for the most part, been doing routine work for the most part. They have moved from routine work on the farm to routine work in the factory and later to scanning barcodes at Walmart. Most people continue to do things that are on some level fundamentally repetitive or routine and predictable. Predictable is maybe the best word to use. That really

means that those types of tasks are going to be susceptible to machine learning going forward. There is a huge impact there.

As I have said, workers, historically, have adjusted. Economists will look at that and give a lot of weight to that, to that historical record. It is interesting and this is kind of an analogy that some people find a bit disturbing, but there is one kind of worker that did not adapt. These workers were completely left behind by technology. They never really found a new role. Those workers, of course, were horses. Now, you can have a very reflexive response to that and say, you know, that is ridiculous. How can you draw a comparison between human beings and horses because, of course, horses are just horses and people are intelligent, people can learn, they can adapt, they can figure out new things to do and horses could not do that? The point here is that the machines that are going to threaten the jobs of workers in the future are not anything like the cars and the trucks and the tractors that displaced horses. The machines that will come after workers in the future are going to be, in a sense, thinking machines. They will have that cognitive capability. They are, in a sense, going to encroach on that fundamental capability that really sets people apart

and makes them different from horses. That is something that is right now, in its infancy, but it is going to get dramatically better.

To continue the analogy of horses, it is interesting to note that before the internal combustion engine came along early in the 20th century and really substituted for or completely displaced horses, there was another disruptive transportation technology before that. That was the railroad. If you were an analytical horse in the 19th century and you saw the advent of the railroads, you would have looked at this and initially, you might have been a bit worried about that because, of course, trains are much faster and more powerful than any horse. Ιt turned out that those fears were unfounded because the railroad turned out to be a complement to horses, not a substitute because you could use trains to haul things long distances, but then when you reached the terminal point, you still had to distribute people or goods locally. For that, you needed more horses. It turned out the more railroads you built, the more horses you need.

One question that we can kind of ask is as we look at these technologies evolving at this point, are we making the mistake of thinking like a horse looking at the railroad? Are we assuming that these technologies are going to sort of remain on the tracks and not evolve into something completely different and completely disruptive that is going to really substitute for people. I think there is a very high risk of that.

I also think that there is obviously a kind of a continuum of possibilities here. At the extreme, you could end up with mass unemployment if really all of the routine jobs go away and if workers are not able to move into other We have to acknowledge that there is what you might areas. think of as a normal distribution of capability among human beings. Not everyone can go and become a robotics engineer or an artificial intelligence research or any of the other jobs that might be relatively safe. I think that we could end up with a future where a lot of people are simply left behind and then a small minority of people, maybe the top ten percent or so, actually find the role that is augmented by artificial intelligence. You will notice that Guru gave some examples there. Typically, the examples of people that will be augmented rather than replaced by AI are doctors and professionals. The question is what are we going to do about all of the average workers in the economy.

Just as a final point, I would go back to the example that Caroline gave and just tell you a bit more

about Google's Deep Mind. If you look at the game of Go, there are two interesting things, at least, about it. One is that at any point in the game, there is essentially an infinite number of moves that you can make. There are actually more possible moves than there are atoms in the universe. What that says is that there is no conceivable supercomputer now or really in the foreseeable future that could ever calculate that or use a brute force technique to solve that game in the way that chess was solved. The second thing to know is if you talk to the best Go players in the world and you put yourself in the position of a computer programmer interviewing them, asking them to explain what they are doing, they really can't articulate it. It is just something that happens in their brain that is kind of intuitive.

Given those two things, you would look at this game of Go and you would say this is something that we should not be able to automate. Economists typically will divide things into two groups. They will say these are the routine things that we can automate and these are the nonroutine things that we can't automate. I think if you look at this game and what was done, how this machine essentially learned to play the game of go and in a short order became superhuman in it, it becomes clear that this dividing line that we have is simply not stable. It is going to move. None of us can really say with any certainty what is going to constitute a routine job that can be automated 10 years from now or 20 years from now. It may be something completely different than what we would imagine today.

I will just end by saying I do think that there is a very high risk here. It is something that we need to think carefully about. I am not at all opposed to AI. I agree with everyone who says we need more of it. But I do think that we are going to have to adapt to it. It may require, eventually, something very radical, something perhaps along the lines of a universal-based income. Thank you.

(Applause)

DR. JOHNSON: Thank you very much. Milind is our last speaker.

Al in Low Resource Communities

MILIND TAMBE

Founding Co-director, Center for Artificial Intelligence in Society(CAIS) University of Southern California tambe@usc.edu

Co-Founder, Avata Intelligence

Presentation of Milind Tambe, 12/12/2016

DR. TAMBE: My name is Milind Tambe. I am from USC. This talk is part of an experiment where I flew 20 hours from India, landed, and I am giving a talk. We will see how well it goes.

The focus of our work that I will talk about today is AI for low resource communities. We have cofounded a Center at USC called USC Center for Artificial Intelligence in Society.

Overview of CAIS Project Areas



Presentation of Milind Tambe, 12/12/2016

The focus of this center is exactly for use of AI for low resource communities, for emerging markets, for people who have not benefited from AI.

The kinds of projects that are of interest to us in the center are things like working with the homeless populations in Los Angeles. One concrete example that I will talk about more a little bit later is with the homeless youth - we have a large number of homeless youth how to spread HIV information among them. Harnessing their social network by using social influence maximization, we can show that AI techniques lead to larger increases in this information being spread. We have conducted realworld pilot tests. I will show you some initial results.

Originally, I was supposed to be on a panel for public safety and security. I won't talk about it much, but this is another focus of the center. So we have a large number of targets to protect and limited security resources. How do you schedule or plan or allocate these limited resources?

Some of this security work we have been conducting for the past ten years uses an AI technique called computational game theory. By solving mass use scale security games, we can generate schedules. For example, for the Federal Air Marshal Service, assigning the Air Marshalls to flights on a randomized basis. That is something that we have done. That is in use for the past seven or eight years. If you have been on an international flight, whether there was an Air Marshall on your flight or not may have been decided by an AI program. Or work that is being used by the Coast Guard, for example, in ports like Boston and New York and Los Angeles. If you have seen these sorts of patrols around the Staten Island Ferry, trying to protect the ferry from somebody ramming a boat

into the ferry, these are AI algorithms at work. These are our algorithms at work. They have radically changed the way the Coast Guard do these patrols.

I will not focus on that. I will focus instead on a third topic:



Overview of CAIS Project Areas

Presentation of Milind Tambe, 12/12/2016

That is AI for protecting endangered wildlife, a concern particularly in emerging market countries. Here, the goal is using machine learning to predict where poachers will attack. If we can locate that, locate snares before they are deployed, then we can save animals.

In fact, there are a very large number of projects that we have ongoing. For example, for working on veterans' mental health, post-traumatic stress disorder. We are also looking at gang violence in Los Angeles. Our faculty are also interested in climate change and impact on low resource communities.

I will now dive a little bit deeper into one of these projects. That is looking at HIV information spread among homeless youth, just so that you have a sense of what is going on. I will then talk just very briefly about wildlife conservation and end with some general remarks.

Here, I am going to show you a video because I want to highlight my Ph.D. student, who did this work, and Eric Reiss, my collaborator. It is a two-minute video. Hopefully, it will all play.

(Video available at:

https://www.youtube.com/watch?v=eWvE7Gvsr9c)

DR. TAMBE: This is part of our submission to our AAAI conference. I have received a lot of criticism from my family members about the acting, the walking and talking. Fortunately, AAAI doesn't grade us on our acting. Hopefully, this video will pass.

I wanted to talk a little bit about how the

HIV Prevention Programs: Using Social Networks to Spread HIV Information



Presentation of Milind Tambe, 12/12/2016

A lot of the links are uncertain. We don't know who is friends with whom. The AI program then first finds out a limited set of peer leaders. We gather information from them about the network, but also ask them about spreading information. We train them about spreading information about HIV and how not to get HIV. With additional information about the network, we train the next set of peer leaders and then the next set of peer leaders.



Challenge: Adaptive selection in Uncertain Network



Challenge: Adaptive selection in Uncertain Network

Presentation of Milind Tambe, 12/12/2016

It is running a campaign where you keep getting additional information improving the selection of peer leaders.

We have written this program. It raises a program called HEALER, an acronym that was very cleverly selected.



Challenge: Adaptive selection in Uncertain Network

Presentation of Milind Tambe, 12/12/2016

We have done, of course, simulation studies looking at homeless youth networks that we got on Venice and Hollywood.

Real Networks - Solution Quality



11

Presentation of Milind Tambe, 12/12/2016

This is showing you that our program, HEALER, performed better in terms of spreading information. The Yaxis there is number of non-peer leader youth to whom information was spread. This is a simulation study.

Then we went out and ran pilot tests.

Pilot Tests: How we generated the network

Recruited:

62 youth for HEALER; 54 for degree centrality (non-AI)

Preliminary network -> HEALER

Bring in 4 youth for training, get more tie data —> HEALER

Bring in 4 youth for training, get more tie data —> HEALER

Bring in 4 youth for training



Presentation of Milind Tambe, 12/12/2016

For example, one where we recruited 62 youth for HEALER, 54 for non-AI program, which basically recruited the most popular youth. This is the way the traditional homeless shelters run this program. They keep recruiting the most popular youth. Here, we are going to use an AI program. The way it works is the preliminary network we get, we feed it into HEALER. It tells us there are the four youth you should bring in. We get more network data. We give it to HEALER. It brings in the next four people. In this way then, we train these people. After a month, we look at how many non-peer leaders got information about HIV.

Percent of non-Peer Leaders Informed Not Informed Not Informed Not Informed Not Informed Al HEALER Non-Al

Results: Pilot Studies

Presentation of Milind Tambe, 12/12/2016

What this is showing you here, on the Y-axis, is percentage of non-peer leaders in the test who got information about HIV from peer leaders. You can see that the AI program led to 70 percent of non-peer leaders being informed whereas the traditional non-AI technique that was used by the homeless shelters led to a much smaller percentage of people being informed about HIV.

We also looked at how many people started getting tested.



Results: Pilot Studies

Presentation of Milind Tambe, 12/12/2016

This is looking at the darker green portion and seeing how many of these started getting tested for HIV. We found that with the AI program, it is higher. We don't completely understand why this is happening, but this is good news. Essentially, now, we are on to running this program with a thousand homeless youth in LA, starting this spring. The IRB (institutional review board) and so forth is moving forward.

At our center, there are a lot of other projects.

Other Applications in Urban Settings

- At CAIS:
 - Substance abuse/addiction and intervention among homeless youth
 - Gang violence in Los Angeles
- From the AI100 report:
 - Machine learning for preventing lead poisoning in children
 - •Predictive models to prioritize houses for lead paint inspection
 - Scheduling and planning to redistribute excess food to food banks

Presentation of Milind Tambe, 12/12/2016

For example, we have received a grant for looking at substance abuse and addiction amongst homeless youth and how to reduce that. As I said, gang violence in Los Angeles.

If you know the AI100 Report, which was recently published, there were many other such topics that were mentioned as work that has gone on in the AI community. Machine learning for predicting which homes should be tested for lead to avoid lead poisoning in children. Scheduling and planning to redistribute excess food. There are many, many possibilities in which AI can be used for low resource communities.

I will talk very briefly about wildlife conservation. We all know the terrible stats about poaching of elephants and rhinos.

PAWS: Poacher Behavior Prediction + Generate Patrols

Murchison Falls National Park, Uganda



Presentation of Milind Tambe, 12/12/2016

This is Uganda, Murchison Falls National Park. Wonderful animals. If you haven't been there, I recommend it. There are also threats to the wildlife. These threats come from these kinds of snares, in fact, thousands of these snares that are placed every year. The question is then from past data, can we predict where the snares are located?

I will move forward quickly. How likely is an attack going to occur on a one kilometer by one kilometer area?



POACHER BEHAVIOR PREDICTION

Presentation of Milind Tambe, 12/12/2016
We can look at different factors that are available to us, range of patrol, frequency, animal density, all sorts of other features. We constructed different programs, tested them. We looked at 2015 data.

POACHER BEHAVIOR PREDICTION



Presentation of Milind Tambe, 12/12/2016

The thing that worked the best is an ensemble of decision trees, essentially, some older technology that seemed to really work very well. The green bar there shows that it was more accurate than other techniques.

That is 2015 data. Train it on 2003-2014 and predict where you would find snares in 2015. We already

have those. We can do a good job. What would happen in 2016? With Wildlife Conservation Society, we ran a test, a one month test.

Real-world Deployment: Results

- Two 3 sq km patrol areas: Predicted hot spots with infrequent patrols
- Trespassing: 19 signs of litter etc
- Snaring: 1 active snare
- Poached Animals: Poached elephant
- Snaring: 1 elephant snare roll
- Snaring: 10 Antelope snares
- Hit rates (per month)
 - Ours outperforms 91% of months

Presentation of Milind Tambe, 12/12/2016

We predicted two areas where we said here is where you will find snares that you do not patrol at present time. This one month test started. We predicted hotspots for infrequent patrols. To give my students additional incentive, I said for every snare that was found, I would buy them a drink.

During the one month period, first, we found lots



Historical Base Hit Rate	Our Hit Rate
Average: 0.73	3

of signs of trespassing, not clear how to translate that into drinks. Then, one active snare - well, that is one drink for all of the students. Then came this terrible news that we found a poached elephant with its tusks cut off. Poachers were active in that area. At least the program had found the right place to go to. Then came the news that they had found a whole row of elephant snares. Poachers were active. They were killing elephants. Now, we found a bunch of snares. They were hidden. Before they were deployed, they were removed, thus potentially saving the lives of some elephants. Then came the news that there were ten antelope snares, at which point, my students said, okay, we can't drink anymore.

We will see - this is work under submission. We will see how our reviewers like this work. Our hit rate is much higher than the base rate.

Let me now come to some key points I want to make.

Towards the Future

- Significant potential: AI for low resource communities, emerging markets
 Direction of AI research in our hands
- Novel research challenges:
 - Fundamental challenges from use-inspired research, e.g., POMDPs
 - Also: Interpretability of AI techniques,
- Designing AI systems in Society:
 - Belmont report principles: Respect for persons, Beneficence, Justice
- Methodological challenges:
 - Encourage interdisciplinary research: measures impact in society

Presentation of Milind Tambe, 12/12/2016

There is significant potential for AI to have impact on low resource communities, on emerging markets, on people who have not benefitted from AI. The direction of AI research depends, in part, in the hands of AI researchers like ourselves. If we push for this direction, then that is where AI will tend to go.

There are some novel research challenges that this area brings forward. These are fundamental research

challenges, but also newer kinds of research challenges that demand that the AI techniques, the results we produce, are interpretable. When we talk to rangers in Uganda and we say, okay, here is where you should go and they say, why, and we say, well, here are some complicated technique we can't translate, but there is where you should go, that doesn't work. It has to be fundamentally interpretable. The reason we use this technique called decision trees is because we could explain what was going on.

Where in our experience have AI techniques succeeded? In building AI systems in society, what we found interesting is that things from the Belmont report principles, things that we have to study if you do human subject experiments, are actually very applicable. Respect for persons. We want to respect the autonomy of the people whom we want to help. Beneficence. They should help rather than harm. And justice, we should try to make AI to be useful or valuable to all parts of society and not focus its attention on a single segment.

For AI audiences, in particular, I feel - one final point, if I can get- is also important. That is increasing interdisciplinary research. If you want AI to actually help low resource communities and so forth, one important aspect is actually measuring its success in the real world. That also means scientists need to be encouraging of this work. This is not easy because interdisciplinary work doesn't get reviewed very well. This is the kind of work where you have to spend months evaluating things in the field. It is a lot of work. Reviewers rather would like to see mathematical theorems and so forth sometimes. This is something that is more of a challenge for our own community to foster this kind of work.

Thank you. That is my presentation.

(Applause)

DR. JOHNSON: Thanks very much to all the panelists. We now have plenty of time for question an answer. Start to formulate those. The three things that I take away from this discussion so far and from - fitting very well with what we heard from Jason, first of all, cognitive computing is coming in a very big way. It is an important new general purpose technology cutting across a lot of different sectors. I am looking forward to having my own cognitive assistant. In fact, maybe I will have several of them and they will argue among each other. That might actually be productive.

Routine is in trouble. Routine tasks are really in trouble. In that sense, while a lot of stuff - a lot of

the things we are talking about today is new, that is not new. We have seen that in some sectors, including in manufacturing, for a long period of time. I am drawn back to this work David Autor has done on job polarization, where some people do very well because they become more productive and they can capture the benefits of that and other people get pressed down very hard because they are going from routine task to routine task and those get eliminated. There are still bank tellers, absolutely, but bank teller used to be a good mainstream, middle class job, not so much anymore, I believe. At least the higher income jobs in banking are doing very different things from bank tellers.

Now, I think Martin puts this very important issue on the table, which is, all right, do we end up with a lot of unemployment? What is the kind of impact? I am a little bit siding with or even quite a bit siding with Jason Furman for the United States. But, I think, Martin because we have some mechanisms that have turned out pretty good about generating new jobs. But for much of the world, that is not true. Much of the world when you lose those routine jobs, you do get mass unemployment. That is true in parts of Europe, for example. That is true in middle income countries. It is massively true in low income countries. The new jobs do not automatically spring up.

I think this general-purpose technology, just like you go to the lowest income part of the world, you fly on a jet aircraft. You drive from the airport in a land cruiser or similar vehicle. That is general purpose technology that is used everywhere across all income levels. I think this is what we are talking about here and the pressure on routine tasks. I would love to be wrong, but we can discuss that.

Third point is, and I think the most encouraging point, and I hope we have time for this today and subsequently, the private sector can and should engage with this technology and these topics somewhat differently. Caroline was talking about education. IBM was talking about different kinds of applications to healthcare, for example. Of course, Miland has got a very broad and inspiring agenda of social impact. I think the criteria that you were laying out - what are we trying to do for people in this, recognizing Martin's points that a lot of people are going to be hit very hard by this and, frankly, I think a lot of people in this country are already extremely disturbed and feel left behind by people like us who have all done very well over recent decades from technology and from trade, by the way.

Okay, let me open it up to the floor, questions, comments, pushback. We have a microphone coming around. Let me recognize the lady on the right of the aisle here to start us off. I encourage you to direct your question and we will try to make sure everybody gets a question.

DR. POLSKI: Margaret Polski with the Krasnow Institute. Thanks very much to all of you. The point has been made - this is, as I understand it, a panel on the economic impacts. Well, policy affects economic behavior and vice-versa. These are - political economies are complex systems that are, themselves, subject to analysis using AI. To the technologists, how can we apply AI as policy analysts and economists to better understand the decision problems we have to make here?

DR. BANAVAR: Let me start with this. Again, I would make a top-level point, which may have been lost in some of the discussion, which is that the analysis of language, as opposed to numbers or pictures, all of that, is very different, more complex, and, in fact, represents one of the big breakthroughs that happened in the Jeopardyplaying machine that we built. I just want to keep that in mind because it comes back to how do we understand knowledge, which usually is encapsulated in language through analysts' reports and publications and so forth

with pictures or diagrams embedded within it.

The direct answer to your question is that some of the most recent work we have done is kind of one step beyond question answering. Question answering was sort of the technology that Watson is best known for, where you have a fact-based question or a fact-oriented question, which there may be several answers to and the machine basically evaluates which one is the best answer based on a set of scoring functions. The next step after that is essentially, an analysis of pros and cons for a given hypothesis.

One of the systems we have built in the last three or four years now, after the question answering system, is something we call a computational argumentation system, which is basically to give a decision maker the ability to evaluate the pros and cons. I didn't have the time to show you an example of that...I have a video and a running system on my laptop here - which has been fed all of Wikipedia. Now, Wikipedia is not a highly reliable source, but you could imagine feeding Medline or you could imagine feeding the Economist archives or New York Times archives or whatever. Once you have such an analysis of the language of that corpus of information, you can ask the machine - you can state a hypothesis like, for example,

immigration is good for the economy - you know, very broad economic kinds of a question - and we have a running version of a machine now, which pulls out all of the information or a lot of the information from the corpus that relates to that hypothesis and puts down some pros and cons of whether that position that you have just taken is positive or negative. It pulls out opinions. It pulls out facts and research and even diagrams and then lays it out in an application on the left pros and the right cons. You can decide, as a person who is trying to make the decision, which things you want to pick and how do you want to combine them and create a first version of a report that you can get started. We have many examples of that being applied right now in industries. I will just use that as an example.

DR. TAMBE: As we engage in these discussions these were very, very interesting presentations - we should be careful to outline what exactly are we talking about when we talk about artificial intelligence. We have here Professor Stuart Russell, author of a very popular book on AI. I am sure he has his own thoughts about what is AI. When I was a Ph.D. student, there would be running battles in the corridors of Carnegie Mellon of exactly what is AI. People would snatch things away from us as AI researchers

to say this is not AI. This is routine automation. We would find it very difficult to claim successes because people were taking them away from us one after another. Now, it is so interesting that all of automation, all of the joblessness that has come from 1970's and so forth, the income divide, all is blamed on us, as AI researchers, when, in fact, we were not responsible for it.

We have to be very careful about what is AI and where do we draw the boundary of just routine automation versus AI. Maybe this is not a conversation purely about AI. It is about all of automation. I just want to be careful about what we are talking about as we move forward.

DR. JOHNSON: I think we will ask Stuart Russell to tell us his thoughts. I think the question is what is new, what is different, what is continuing, in what ways is this going beyond what we have seen already?

PARTICIPANT: I enjoyed Jason's remarks. They were very reassuring in some ways. If you ask why have AI and robotics not had a huge impact on employment, it is because AI and robotic systems are or have been blind, deaf, and illiterate. You ask, okay, well, is it that surprising that they are not taking a lot of jobs away from humans when they are blind, deaf, and illiterate? No, of course not. Those limitations are going away very rapidly.

I recently was talking with Andy MacAfee, who has written about this. He says, in his book, tasks involving a lot of hand-eye coordination, sensory motor skills are unlikely to go away in the near future because, you know, robots can't do them, particularly in unstructured environments. I showed him a video of our robot at Berkeley just taking a pile of towels and picking them up one by one and folding them up. He said, oh - and then he used a swear word - that is another 500 million jobs gone.

I think we are going to see quite rapid impact as blindness, deafness, and illiteracy go away. I agree completely with Martin that we have been using humans as machines. All of the jobs where we have done that will go away. The argument that, oh, well, we have had these revolutions before and we always found new things to do, well, yes. Once you have replaced all of the physical skills of humans and you have replaced all of the mental skills of almost all humans, then what is the next thing that humans are going to do? I think you have to assume that the next thing that humans are going to do is be human, which is hard to replace.

Most people are not going to be in manufacturing or the routine services. We may still be writing screenplays and doing this kind of thing. Most people are

going to be in one on one human services. Those professions, right now, range from psychiatrist all the way down to massage therapist, across the whole income scale and education scale. They all involve interpersonal skills. They are almost all self-employed. They are not STEM educated. Teaching everyone about AI so that they understand why they don't have a job is probably not the best thing for the future of the economy. I think we have to look 20 or 30 years ahead and be realistic about what people are going to be doing and what kinds of skills they will need to do that. To me, it does not look like lots and lots of data scientists.

DR. JOHNSON: I don't think it was a question, but it was a powerful statement. I think everyone on the panel should have a chance to respond. In the interest in mixing it up, I will take Martin and Caroline and then if you also want to chip in. Martin?

MR. FORD: That is pretty much the argument I have made. I agree with all of that. It is true we talk a lot about the cognitive aspect of this, but definitely the fact that machines are becoming more dexterous and mobile - I mean there are huge numbers of jobs that are protected for people now not because they require some high level intellectual capability, but because they require that hand-eye coordination. The machines are definitely moving into that area.

The biggest disruption there could well be in countries like China and in developing countries. It will mean - it is really important to keep a global focus on this. It is not just going to happen in developed economies, but across the world, the path to prosperity for developing nations has always been build factories, create huge numbers of jobs for unskilled workers. I think that path is going to evaporate. It is just a global problem that we really have to think about.

MS. ATKINSON: Perhaps unsurprisingly, I have a different view. I have noticed moving from the world of economists and policymakers to a world of people who know about technology and are advancing technology that the intersection of these two worlds is very narrow. I think that leads to these different conclusions. It is important to think what are we losing by not having the intersections closer.

Now, the economists are failing to understand - I think that I would hear from you - how rapidly and dramatically AI is going to impact on what is possible for non-humans or machines to do. What is missing from the technologist's view of a future without jobs is the

knowledge about how governments can actually have a huge impact on what governments and markets and their interaction on what is the level of demand, how is it that people stay employed without having people pick particular jobs. We have just seen over and over again, as Jason laid out, that other things turn up for people to do.

I think that - I want then just to pick up on what Martin says about the global view. I think that is really important to recognize. China already doesn't do the low skill manufacture. They do assembly. They could have had a big concern - well, we are not going to be able to employ all of these people because Vietnam and Indonesia and others are doing the low skill manufacturing, but in fact, it was possible to find other ways for people in China to be employed. People in Japan and increasingly in China, are very worried about who is going to do all of the work to support them in old age because of the demography.

I think that I am much more firmly on Jason's side. Even though I can't imagine and many of you will be much better than me at imagining - and many of my company, what are the tasks that - and what are the skills that machines and computers can have, I think that there is so much need and want in the world that it is very hard for me to imagine that all of those needs and wants get met just

by the application of machines. As they are increasingly met - you know, there is nothing God-given about a 40-hour week that we enjoy in the advanced world. Why shouldn't there be better factory conditions, better improvements in the lives of people in emerging economies?

To me, the real difficulty is what Simon referred to, which is how do we - how do policymakers with the private sector, as well, address the uncertainties and difficulties of people faced with changes in the external environment? How do we promote the ability to move? How do we guard against increasing inequality, which is corrosive in society? How do we make sure that there are still institutions? Because your point, professor, about more self-employed people does raise the question a lot of people have found a power and dignity in joining with others in similar work. That is obviously harder if work is distributed.

I think all of those social science problems go back to the issue raised by you. We need a lot of humans thinking about that, doing political work to address it. Thanks.

DR. BANAVAR: I think I will just add one more thing about education. The main point, really, is not about STEM. I agree with you. It is really about skill-

based education and probably, you know, sort of the structure we have in vocational education. It is more skill-based than what we have traditionally.

Perhaps I think the move towards that kind of skill-based education should accelerate, not just for teaching how to recognize when a task can be done by a machine versus when it cannot, which is an important skill for anybody to have and I tell my kids that all the time. So, think about every task you do. Don't do it when you know a machine can do it for you. Look for when you can. Go use Wolfram when you have to solve an equation - that kind of thing.

I think there is a set of skills that I believe are required for creative problem-solving. That is not something that I have seen machines be able to do right now. As a technologist, we can look towards - probably not in the near future - that you can do management types of jobs or entrepreneurial jobs and those kinds of creative problem solving jobs. That is probably one of the frontiers that people don't understand how to build skills for. But I believe the human-level jobs, which use much more interpersonal skills and so forth are definitely something that we need to teach people a lot more about. I think that should be the focus of skill-based education, in

my view.

MR. NITZE: Hi, Bill Nitze, again. I like the phrase if you can't beat them, join them. What I mean by that is a mean for a general intelligence is 100. Everybody in this room is at least one, maybe two standard deviations above the mean. Let's take mirror image, one or two standard deviations below the mean. You have trouble with the railroad schedule. That leads me to the question, how about some enhancement in our mentation? I would like an upgrade. Shouldn't we be thinking of enhancing human intelligence by incorporating this technology into us so that we keep pace and are not obsoleted? I know there are a number of people who have talked about this, but I haven't heard mention of it yet this morning.

DR. JOHNSON: We do have a panel on ethical issues coming up. Let's not stray too far.

DR. BANAVAR: Just to clarify the question though, you are thinking about not just an assistant - a machine internalized, like embedded.

MR. FORD: I will take a shot. That is a concept that a lot of people in Silicon Valley talk about. Ray Kurzweil and his whole crew of singularians are very into that idea of augmentation. I think that to some extent that will happen. It is hard for me to imagine that that technology is going to become available to everyone at every level of our society any time soon. It is going to be the billionaires that will have access to that first. That may actually increase inequality rather than decreasing it. Perhaps that is something of a solution in the long-run, but I don't think it is something that is going to address our issue in a meaningful timeframe.

DR. TAMBE: I guess I wanted to quickly say the decision aids that we build that Guru talked about are a step towards enhancing cognition. That is certainly the case.

I also wanted to say very briefly about the developing world - so I heard stories from the first IBM engineers in India, who were trying to sneak in a computer into a factory. There were workers standing outside saying they will absolutely not allow this IBM computer to go in. During a festive system late at night, they put the computer in and it started working. This has been a path to prosperity, I would argue, this computational science. Similarly, I feel AI - there should be an AI surge in the developing world, focused, targeted surge that would similarly provide significant benefits. That would be the way to move forward in my view.

DR. PEZESHK: This is Aria Pezeshk from the FDA.

I actually work on deep learning for medical imaging. Regarding the point you made about education, I would say that is going to be very difficult. Even somebody who has worked with machine learning the last couple of years, it has completely shifted. Somebody who worked with neural network would have been made fun of in graduate school. Now, all of a sudden, deep learning comes along and everybody is using basically, neural networks. To the point you made about education and skill, how do you decide what you want to do if the landscape is changing so fast?

DR. BANAVAR: I do not mean to say that everybody needs to be taught machine learning.

MR. SAHINER: That is not the point. Let's say you want to go into accounting. Accounting might disappear four years from now.

DR. BANAVAR: When you think about medical education, I believe that rather than sticking with a stethoscope, from I don't know how many decades ago, I think medical education should be fundamentally changed for doctors and physicians to use data scopes that are pretty much like - you know, you go to the bedside or you go to a lab and you have a data scope that gets you the right context of the right person or the right situation so you can make your next decision. What exactly is a data scope and how, exactly, it will evolve, I think that is going to obviously change every few years and so forth. Making that shift into a data and evidence-based and a rapidly changing dynamic situational awareness is the kind of skill that I think doctors should be taught in the future as opposed to having a narrow band, listen to a person's body kind of a training.

That is an example of the kind of thing I am talking about. I think we need to rethink how we educate every professional in every industry.

DR. JOHNSON: I think it is a brilliant question if I might interject. This course, which we are running at MIT, is attempting to experiment with various dimensions. Johnathan Roland, who is working with me, sitting right behind you - and we are here today, in part, to get these ideas. We want to do it in an open way. We would like to share that with everyone.

I think how you teach accounting, for example, up to this point, we should take the same sort of idea and make that kind of work more productive. I don't know if that is STEM or something different, to Stuart Russell's point, but I think that is a really good question.

I also think the private sector, such as people represent at this table, universities and corporate, need

to take a lead on this because you are bringing us this technology at an ever-increasing pace. It is going to transform many things. I think you are saying or what I am hearing you saying is that you agree that you share responsibility with the rest of us for thinking about how we get social benefits and how everyone can gain from that as opposed to just doing it and let the consequences fall out where they may. I think that route is a little bit what has gotten us into trouble so far.

MS. ATKINSON: I think that is right. Speaking just for Google, we do spend quite a lot on - we do a lot of work on digital skills. Now, there was the question about whether those are the ones that are useful - will be useful. Right now, we know that, never mind AI, it is very useful, for example, for small businesses to learn how to have a website, to learn how to advertise on the internet, find new customers that way. It is also very useful to for students in schools and so on to have access to the internet more widely. There are educational programs and ways that companies like ours can help.

I also wonder, though, and this is something we are always wrestling with - there are lots of social issues where, say, employees of the company feel are very important to address. One of the things, if you just look

at education, look at all of us in this room. It is a heavily white male audience here. That is also an issue for how we spread education and learning and how we prevent AI from becoming a technology that is largely, initially, benefitting this kind of audience.

I think that - again, maybe it is my background as a policymaker. I do think that government has an enormous role to play. Politics and politicians have a role to play in deciding how best to do this education and to collaborate. Jason pointed out that business investment has remained high, government investment, especially in basic research, which doesn't have an immediate pay-off that attracts business, has fallen. That is a sort of public choice. If we have so much stuff and so many abilities, why aren't we asking our government to do more basic research, for example? We talked a little bit about - I am well aware of the difficulties of getting policy changes, as Jason said, but I think that we have to think about would it make sense for governments, the public sector and the private sector, universities, all of these different people to play their part. Thanks.

DR. JOHNSON: We have about 12 minutes left. Try to get as many questions and voices in. Gentleman on the aisle here? MR. FINKELSTEIN: I think the next big thing is going to be the advent around 2020 or so of commercially available, autonomous intelligent vehicles, driverless cars, trucks, buses, and so on. Within the decade after that first introduction, a little beyond - more than three million drivers in the U.S. alone will lose their jobs. There will be all kinds of disruptive and transformative changes not only in the industry, the automotive industry, but high-tech industry and society, in general. The opportunities for new entrepreneurial activity, a lot of which are difficult to predict now, there will be military implications, law enforcement, national security implications.

I would like to hear any comments from the panel, since this is very near future. After that, there will be subsequent waves over the decades to the end of the century with that as a basis for achieving autonomous intelligence, only in other forms of robots as well, humanoids, and so on. Anyway, I would like to hear some comments.

DR. JOHNSON: That is a great question. I think the big one for today. Why don't we do a round, just come down the panel? Milind, driverless vehicles impact, 2020 on, in about 20 seconds or less.

DR. TAMBE: Since the question was more general

about all kinds of autonomous things, I feel as AI researchers, it is our responsibility and for people involved in AI, to really focus AI's attention towards people who have not benefited from AI, for the low resource communities and so forth. That is obviously the focus of our center. That is an imperative that we need to undertake in order to mitigate some of the issues that you are bringing up that could occur if we let AI sort of take its own course.

MR. FORD: Right, I agree, it is going to be a big disruption. As you say, vehicles, driverless vehicles, will be millions of jobs. Of course, it won't just be that. It is many other areas that will happen more or less, simultaneously. You may have seen that Amazon just introduced a prototype store that doesn't have any cashiers at all. You just walk in and grab it and walk out.

All of that is going to happen at once. I am concerned about the impact on society and on politics. I mean we have just had an election that I think produced an unexpected result. I think we can expect a lot more of that in the future as more and more people feel that they are really left behind.

DR. BANAVAR: My quick reaction is that, yes, there is going to be a period of transition. I don't know

how long that is going to be. In a future generation, that job is not going to be on the list of jobs that people are going to look for, obviously. During the period when the existing job holders are displaced or transition, I think retraining is the question. I think that is what we need to focus on. I don't know what - I am not an economist or an educator. That is the question that we should be asking ourselves.

MS. ATKINSON: Yes, I agree about retraining. I also think that the kinds of policy interventions that Jason talked about with wage insurance and so on are needed.

I remember when I was growing up and Thatcher was there and closing down coal mines. That was a huge problem. It was partly a huge problem because the overall economy was running at well under full employment. There were no alternative jobs. There had been a similar sort of shift away from railroads in the 60's, I think it was, when Beaching closed down all the railroads, which caused some problems for consumers because their local stations stopped servicing them. However, the economy was booming and the people who had been working on the railroads found other jobs. They weren't directed to them, but those other jobs were available.

I do think this notion that there is no ability to transition is just wrong. People don't - what was that data? It was like six million jobs are created and five and a half million jobs are lost. There is a huge amount of churn. It is not just that you will have three million people now working as drivers no longer working as drivers. There will be an evolution. These things will spread in different ways. Maybe people won't have to drive for so long in order to get the same level of income and access to consumption. There will be fewer accidents and more people to service. That could all be fine.

DR. JOHNSON: Five minutes left. Questions.

MR. FORSCEY: Thank you. My name is David. I am with the National Governor's Association. I think so far we have had this conversation very much at the national level. As a lot of us know, a lot of economic policy is determined at the state and local level. I think we may end up seeing a mismatch between what some people in Washington might want and what some people in the states might want. I wondered to what extent researchers are working with state and local governments now and how they see that role playing out in the future.

> DR. JOHNSON: This time we will go down this way. MS. ATKINSON: Very quickly and not as a

researcher, obviously, with driverless cars in the United States, state and local governments set a lot of regulations. People are working very intensively there. I don't mean to minimize the shifts that can occur and the disruption that that can cause to individuals, but I think that one should search for the solutions in government policy and in research and not necessarily just in technology. It is bringing the two together. Thanks.

DR. BANAVAR: I live and work in New York. I am very highly engaged with the New York State government. Just last week, in fact, I had a conversation with the New York State science and technology body that advises the governor. Actually, the same exact kinds of questions that we are talking about over here are being asked at the state level. I have also served in Governor Cuomo's state resiliency program after Hurricane Sandy and so forth. Yes, I think that needs to happen much more broadly across many other states. I would encourage my team members to do the same thing across all of the states that we do business in.

MR. FORD: I think there are some what you might of as nearer-term things that could be done at the state level. One obvious thing, I think, would be more investment in community colleges because those are really

the vehicles that are going to enable people to retrain.

In the longer-run, if the concerns that I have expressed really develop and we need something like a basic income, a guaranteed income, then I think that has to be done at the national level. There is a mix of policies that are appropriate to different levels, I think.

DR. TAMBE: Our work with the spreading of information about HIV among homeless youth is funded by the state of California. We are very much engaged. We want to work closely with our city in Los Angeles and so forth. It is a two-way street, too. I guess from the government side, if there is more emphasis, more encouragement of work on AI, that would benefit marginalized sections and so forth. That is where the research will go. In part, it is a two-way communication, I feel.

DR. JOHNSON: Let us take our last question.

MR. ZITNIK: My name is Gerard Zitnik. I am a fellow at the Department of Justice. The question I have is not to do with unemployment that may occur, but rather the possibility of underemployment or a race to the bottom that may occur. I am not just talking about in the STEM fields, but in general, it seems like, oh, it is a good idea. It is nice that we would have a more educated workforce if we had these lower income jobs lost and we

retrain people. Trade unions, for example, have a long, long history of putting regulations into place, hours that you have to be able to get in there to make sure their wages stay high.

If we retrain all of these people in higher forms of education, what have you, we could see a race to the bottom because you are going to flood the market now with all of these people with higher degrees. We see it right now. Look at the rise of adjuncts that has occurred. If new funding and research and demand does not increase, but you keep these people getting these jobs, you could see underemployment. We are seeing it right now in higher education.

I just wanted to know what the panel's view was on this if you thought that these markets are just going to continue to expand or if you see that - it was nice to see that people were talking about, okay, we don't just need STEM jobs. I don't really hear the talking heads on TV really talking about that. They are really throwing STEM out there. There isn't a rise of art schools. I just wanted to know what you guys thought about the future of employment with increasing skilled labor and increasing higher education.

DR. JOHNSON: I think this is a good way to end,

the future of work. We talked about the low-end piece, the driverless cars piece, but what about pressure at the higher end and over production or certain skills and under employment of those skills. Let's do 30 seconds to tie it all together for us.

MS. ATKINSON: Again, my basic view is that we need to make sure that there is still sufficient level of demand and that people have access to resources. I think the problem with the universal basic income is that it is indiscriminate in that it would give everybody access to a certain income. If we were to become incredibly wealthy, maybe. Before that, it is important to target support.

On the point of the higher educated being underemployed, I guess it depends whether you mean that they want to do more work than they are - than is being demanded or whether they are not getting as much income. Clearly, there will be shifts, although much lesser shifts in the high skilled areas. I don't think that is going to be the concern, such a concern as in lower-skilled areas.

The same sort of policies are the ones that we would need to be looking at, what do people need to be retrained in, educated in, supported in the transition. That involves state and local policies as well as national policies and international policies. Thanks.

DR. BANAVAR: I am not an economist but I can tell you how large companies are transforming higher skilled workers, like IBM as an example and probably many other mid-size companies as well, definitely the Fortune 500 or Fortune 1,000 companies. At IBM, we have something like a few tens of thousands of sales people, as an example. These folks are now using AI tools to do research on customers that they call. I am involved in building those tools. We are educating them on using those tools. Now, that is not going to take away - I don't think - sales jobs. I think it is going to make them more effective in doing what they do. They probably would have spent a huge amount of time doing client research before making a client call, but they can probably do a lot of research on their mobile phone while they are driving to the client.

I could go job by job and say how this is transforming each of the jobs and making things more effective. My view is the demand is going to be for higher effectiveness. Therefore, they need to be trained incrementally towards that.

MR. FORD: In general, I agree with your concern. I think that I am very skeptical of the idea that we can educate ourselves out of this. In part, that is because clearly, the technology is beginning to impact higher

skilled jobs. It is going to be an enormous impact on more routine white collar jobs, for example, and entry level jobs are going to be especially impacted. I think we have all heard stories already of college graduates who are working at Starbucks. I would expect to see more of that, especially if you keep pushing more and more people to go to college. There just isn't any guarantee that the work is going to be created for all of those people.

DR. TAMBE: There is a lot of uncertainty and there is a lot of - we are having all of these very interesting conversations about the future of AI. We have made our research bet on where we need to go with AI, which is focusing on certain parts of the population. At the back of our mind as researchers is also what happened the last time that AI was in demand and there was all of the hype. At our conferences, all of the tech companies would show up with free food, free movies, free everything and overnight that disappeared when AI winter came in. There is that additional uncertainty about, you know, when we have all of these conversations, are we talking way ahead of where AI is going to go? Perhaps not. Perhaps this time it is different, but perhaps this time it is the same. I just want to be careful about the hype cycle and what is going on with that.

DR. JOHNSON: In conclusion, I can readily imagine that in 10 or 20 or maybe it is 100 years, this kind of conference will be - there will be no humans. It will be artificial intelligence having these arguments. However, in the meantime, we are all human so there is a coffee break. 15 minutes back there please.

(Break)

Agenda Item: Panel 2: Ethical, Legal, and Regulatory Considerations: considerations of privacy, discrimination, and international norms

MR. GREELY: Good morning. As one of the West Coast people here, I can say that finally - at this point, it is finally beginning to become morning for me instead of late night. My name is Hank Greely. I am moderating this panel. I am here to some extent under false pretenses. I am a carbon-based person and not a silicon-based person. I work on ethical, legal, and social implications in the biosciences, but I also am a member of the Committee on Science, Technology, and Law, which is one of the cosponsors of this event. In that role I do everything that Anne-Marie Mazza tells me to do, which generally has worked out fairly well for me - so far.

In my world, ethical, legal, and social implications of the biosciences certainly are a huge set of

issues. When I look at AI and robotics, both the little bit I already knew about it and the great addition to my knowledge that is coming from today's meeting, it seems to me that the issues are both similar and different. They are issues of privacy, of prediction, of discrimination, deep issues of what it means to be human. On the other hand, the differences are interesting. On the robotics side, people worry about employment. No one ever actually seems to worry that biosciences is going to cure all of our health needs and put the ten percent of the population that works in the healthcare industry out of jobs. Would that we were so lucky.

The issues are similar, but different. One deep way in which they are similar is that there are a lot of them. With four panelists, we can only really begin to scratch the surface. You are hearing four aspects of ethical, legal, and social issues raised from robotics and artificial intelligence, but only four out of what I think will probably ultimately be 15, 20, or 25.

We are deviating a little bit from the order in the program, which you may have noticed is purely alphabetical, although we are not deviating very much. We will start with Missy Cummings, then Robin Feldman, then Laurel Riek, and end with Deirdre Mulligan. I have asked
all of them to speak for ten minutes. I have promised them that I will put the most obnoxious alarm sound on my iPhone to go off at ten minutes. I hope we can hold to ten minutes, both so that we can have plenty of time for discussion and so we can get you all to lunch on time. I will say no more and turn the podium over to Missy.

DR. CUMMINGS: Thank you, Hank. I am Professor Missy Cummings. I head the Duke Humans and Autonomy Laboratory. I am also the head of Duke Robotics. My lab looks at the intersection of humans and technology, both from a very low level design aspect all the way to public policy impact.

We have a very fun project right now trying to work on the development of your own personal flying drone that flies you around, like to the grocery store. It is not coming any time soon, but in the early stages. We do surgical robotic work. I am also working on developing this new project for explainable AI, specifically for designers and users of machine-learning algorithms. We find that this is a big area of confusion, particularly for my own students.

I am not really going to talk to you about any of that today. I am really going to focus on transportation issues, which are really more broadly classified as safety-

critical systems.

I call myself a former heavy equipment operator. In my bio, you will see that my one 15-minute claim to fame is that I was one of the Navy's first female fighter pilots. I flew F-18s and dropped bombs at one point in my life. What is interesting about that is it is because of automation and AI that I transitioned out. Not 100 percent, but it was definitely a contributor. Prior to flying F-18s, I flew A4s, where I did all the flying. In F-18s, I was not even allowed to touch the controls on a take-off. I had to show everyone on the aircraft carrier that my hands were not touching the controls because a person in a F-18 launching from an aircraft carrier can only do bad things if you touch anything. At one point in time, I was the best of the best, but I wasn't better than a computer and I was relegated to the level of a five-yearold.

It was at that time in my life that I realized something is changing. Things are changing. I need to look around. I need to probably change my career path. I did. I went back and got my Ph.D. I was ten years at MIT and now I am down in North Carolina.

That has actually set the stage for my career. I was one of the earlier researchers in drones, and worked a

lot on military drones. Then I started working on civilian drones, then Jeff Bezos commercialized drones. That made my research field kind of obsolete, since it has transitioned. Then I moved over to driverless cars. In that timeframe, I have also been working with the Federal Railroad Administration and other railroad companies like Alstom and CSX, to look at automation of rail.

I have been seeing a very strange development. Since I get to sit and work across all of the major national transportation institutions, such as the National Highway Transportation Safety Administration, FAA, Federal Railroad Administration, we are seeing something that is very strange developing with the advancements of AI in our society.

On the scale of what is the most complex transportation field to automate, railroads are the least complicated. Our country should be embarrassed. The reality is that the United States is woefully behind in its railroad automation implementation. We have embarrassingly too many accidents that lead to deaths because we haven't implemented positive train control, which has been mandated by Congress, but seems to be a very fluid regulation - a fluid law. Really, the railroad agencies can implement it whenever they want, despite the fact that we theoretically

have been mandated to do so. There are clearly some regulatory implications of that.

Again, the easiest one that could be automated forget AI, just plain automation, is railroads. Should have been. Hasn't been. Is in other countries. Is in other third world countries.

The next most complicated technology to automate may come as a surprise to you, and that is drones. Drones are actually a little bit more difficult to automate than trains. They do have a third dimension, which is why I think most people think that it is more difficult to automate. Compared to cars, which is where I am going next, drones do not have the same obstacle density field, our sensors are better, they are far more mature. We have had planes - not just F-18s, but commercial aircraft have been landing themselves quite reliably for over 30 years. There is no question that planes, in general, should be fully automated. Even today, airline pilots of any length of flight in an Airbus plane are only touching the controls three minutes out of every flight. They are only touching it just to give them a security blanket. They don't need to touch it. That is how advanced the aviation is.

Then we come into driverless cars. Driverless cars are - and I have testified in front of the Senate

Commerce Committee about this if you would like to read my testimony. I am a big fan. I am a roboticist. I am here to put technology in the hands of everyday people. We aren't there yet. We aren't there yet for driverless cars to put them in widespread implementation because the sensors are very immature. The Joshua Brown fatality in Florida could not be a better illustration of this. We, as a community, do not know how to do testing and evaluation of stochastic systems like what you find in driverless cars. For the most part, I would say it is an incredibly immature field.

Yet if you think about that three-tier complexity that I have given you, trains, drones, driverless cars, the first ones that you are going to see in this country in your hands are driverless cars, maybe then drones, and whenever the railroad companies decide that they want to comply with the law, then we will get automation in rail. It is a very curious development and as a result, at Duke we are trying to develop a whole new program, an integrative robotics degree program that is both robotics and public policy, where public policy liberal arts students work hand in hand with engineering students to get a joint degree together because we feel this is such a clear problem. How is it possible that we are putting technology on the road that is beheading people like the Joshua Brown incident, but we cannot do the simplest forms of automation to prevent trains from derailing when we know, for a fact, the railroad engineers are either sleeping or playing on their iPad? This is a reality that we are just not coming to grips with. Of course, there are huge industrial implications. The money that is behind driverless cars, the sucking sound that you hear from the driverless car industry soaking up every available live body. If you can spell autonomous car right now, you can be hired in autonomous cars right now.

You are laughing and you think I am kidding. I am not kidding. I was on a DARPA project for a robot copilot. It is the ALIAS project. You can look at the Google videos and see how cool that is. This project came to a halt a couple of times over two years because in one case, Uber completely bought out our actuation team. We got a new actuation team. A few months later, I think it was Comma.ai bought the new actuation team. I actually have been offered a job in the autonomous car industry, despite the fact that I am the human designer, I am not the car designer, because people are so desperate to get those kinds of engineers because - I call it the space race of

technology right now. It is 100 percent in driverless cars. Amazon Robotics, which is one of the coolest companies that you could work for, is now losing people to the driverless car community.

When we have this kind of cornering of the market, of the intellectual capital, that raises problems. It raises problems because we don't have intellectual capital inside the government to regulate this. FAA and NHTSA both suffer from a very, very serious problem in that they have no one inside these organizations that understand how autonomous cars and/or drones work. You think I am being critical. I worked for the government for two years. I ran a robotic helicopter program for two years. I worked closely with these agencies. I feel quite sure that - in fact, they, themselves - even NHTSA has admitted we don't have people on staff that really know what we are doing in this area. Universities can't put them out fast enough. When they do get put out, they get bought up by the car companies.

I think we are at this very important point in society. It is a tipping point. It is a tipping point of how are we going to regulate these technologies. We are so uneven. It is so unbalanced right now. It does put public safety at risk and a host of other issues. The Department

of Defense doesn't have a good robotics outfit right now. What does that say about the state of the world that we are in right now?

With that, I know Hank is giving me the hook. I could talk all day, but I will defer now to Robin.

(Applause)

MR. GREELY: You even had 15 seconds left. I should note that the full biographies - not full, but partial biographies of all the speakers are in your materials, which is why I am not introducing any of them.

MS. FELDMAN: Thank you, Hank. I am Robin Feldman, Professor of Law at the University of California Hastings. My role this morning is to talk about intellectual property as it relates to artificial intelligence. It is a wonderful combination of the technical and also the soaring theoretical.

I will start with the technical. When we speak of intellectual property for artificial intelligence, we are really speaking on two levels. The first is intellectual property in the AI program itself, which could be embedded in software or embedded in the hardware. More important, we are also talking about rights related to that which is produced by AI, whether the thing produced is a video game display created while the user is playing, a process like the direction a car takes, a disease treatment, or an aggregation of data. These are primarily copyright and patent questions. If there is time, I can touch on a couple of trade secret issues.

In the copyright realm, software is protected by copyright, but the American legal system has only brushed against issues related to copyright protection for those things that are created by software. To touch on just one of these issues in copyright, the copyright office has declared that only things created by humans are copyrightable. In addition, copyright case law, unsurprisingly, frames relevant copyright discussions in terms of human creativity. Neither notion, however, has been legally tested or at least not to any great extent.

We do have a couple of appellate court cases finding that images created during interactive video gaming are, themselves, copyrightable.

We also have a marvelous series of cases in which people claim to have written books completely controlled or dictated by supernatural forces. Someone later argues that if the books were written by supernatural forces, the human authors, therefore, have no copyright. The courts in these cases side with rights by the humans, discounting rights by non-human forces, and these cases have been cited by humans

claiming rights in artificial intelligence. I suspect that case law, however, merely reflects a healthy dose of skepticism about the supposed supernatural participants, rather than any logic that would relate to the question of who owns the products of artificial intelligence.

Although the copyright office rules cast a shadow on copyright for things created with artificial intelligence, some other countries have more favorable rules, including the UK, where the entity that directs the AI computer gets copyright in the generated product. It is something like the U.S. "work for hire" doctrine. This is an opportune time, by the way, to contemplate copyright in artificial intelligence in the U.S. Now, I recognize our keynote speaker said we aren't to think in terms of whether things might be different this time, but I am about to say it might be different this time.

Specifically, there is a restatement of copyright law underway, the first time we have ever done a copyright restatement. I have the privilege of being one of the nine academic advisors to the project, and the restatement is likely to consider a number of rules that relate to AI. Similarly, there are signs that Congress may be considering an overhaul of the Copyright Act in the next term. If so, we could try, for once, to be ahead of the curve on a legal

issue.

For patent law, the issues are even more complex and the barriers for obtaining rights are even greater. At one time and for decades, one could obtain a software patent or other process patent by claiming little more than the result. In other words, in very reductionist form, one could claim a program for allowing a car to speed up or slow down based on the speed of the car in front of it, but not offer many more details than that in the patent, itself. Those days are gone. In a quartet of cases, the Supreme Court has cut back on the Federal Circuit's doctrine in this area, culminating in the 2014 Alice case, which drastically limited software patents by setting up a two-step test.

In fact, the only thing the Supreme Court justices seem to be able to agree on these days is that they disagree with what the Federal Circuit says - whatever the Federal Circuit says - although it seems to be a message the Federal Circuit, itself, is reluctant to take to heart. This fall, the Federal Circuit pushed back yet again. In a series of cases on software patents, the Federal Circuit has eased what I call the two-step tango. However, it is hard to imagine that the Supreme Court will greet these decisions any more warmly than they have

greeted the Federal Circuit's decisions in the past.

In short, it is tougher to claim software patents these days, any software patents at all, let alone patents on things created by software, even if one could get past the notion of a computer as the creator.

There is another problem in patenting and licensing of artificial software, and it is one already encountered in the life sciences arena with research tools. For many AI inventions, what you really want to control is whatever is created by the AI program. That is where the money is. However, if you have a patent on something and you license it in a way that asks for a return on what the program created, that is called a reach-through royalty. A reach-through royalty could be considered misuse of your patent, trying to control something beyond what you have patented.

Now, the Federal Circuit all but ended the doctrine of patent misuse in 1992 in a case called *Mallinckrodt*. The Supreme Court has been slowly rejecting that case piece by piece. They granted *cert*. in another aspect of it a few weeks ago. I would not want to bet the farm on any aspect of *Mallinckrodt* surviving over the next few years, meaning patent holders should think about the return of those doctrines and the limitations for

licensing.

With both copyright and patent, there is also the question of the collective nature of the creativity involved in artificial intelligence. In many cases, AI programs generate results by culling data from large numbers of people. Compare that to the way we discuss issues about rights that have become standards. One of the many reasons we worry about granting protections for standards is that the creativity may rest not with the single person claiming the rights, but with the combined actions of many people across society. Similarly, if artificial intelligence programs are deriving their creative results through the collective decisions of numerous human beings, should that creativity be solely attributable to the program? Is this a good fit for our intellectual property system?

I have described challenges for protecting artificial intelligence and its progeny through copyright and patent. The reality, however, is that with most artificial intelligent products today, control operates through contractual agreements, in which a user agrees to assign rights and relinquish claims for the benefit of the program's operator or creator. For the legal mavens in this room, that raises federalism questions, which are always devilishly complex. Specifically, copyright and patent are federal statutes. Contracts are creatures of state law. If state law enforced contracts are allowed to create ownership rights when the federal scheme has declined to protect those rights, are states undermining federal law in a way that results in preemption? One could spend hours parsing that question without even beginning to guess how the Supreme Court would handle it.

Having gone through a whirlwind tour of the technical issues, I would like to close with a few more abstract and theoretical issues. Having listened to the conversation, I suspect that for many people in the room, the theoretical issues are more interesting than the technical legal aspects.

Relying on an individual's ability to negotiate satisfactory agreements and protections in what are essentially vast, far-flung, and fluid communities may be unrealistic. It may also reflect a somewhat distorted view of the nature of legal commons and one that is historically naïve. Although modern Americans exhibit a great nostalgia for the New England town hall meeting, the societies themselves were quite oppressive, imposing strict community norms in a harsh and unforgiving fashion. Communities are not always that respectful of the individual. They can be remarkably intrusive. While the notion of a freely flowing system has great romantic appeal, free flow is only a hair's breadth away from free for all, a game in which the powerful tend to triumph.

I am also captivated by the life science aspects of these questions as we edge closer to creating and implanting within humans, things that are akin to artificial and replicating cells, organs, elements of life, genetic repair. Who controls these? After all, they are literally embedded in us, but if a parasite has embedded itself under my skin, does that make it mine?

As we try to answer questions like these, there is an increasing fluidity of boundaries between individuals and the companies that provide the products and services with which those individuals interact. From the vantage point of my generation, I have an instinct to see this spilling across boundaries as worrisome. I find that the new generation, however, seems to be able to travel more readily across such boundaries and to conceptualize them differently. It could signal blindness, but on the other hand, it could signal an openness to thinking in more creative and less stratified terms in a way that will lead us into a better future.

Finally, for my last ten seconds, I would like to

say that only once in my career have I ever been on a panel entirely of women, and that was a panel on women in intellectual property. I would like to thank Hank and Ryan for today's all-woman panel.

(Applause)

MR. GREELY: I actually thought the notable thing about this panel was that four out of five of us are from California.

DR. RIEK: Hi, I am Laurel Riek from University of California San Diego. I am a roboticist and a computer scientist. I work on building robots that are able to sense, understand, and work with people. I work with older adults and people with disabilities, as well as the clinical workforce, which is the primary space that my work consists in.

What I want to talk to you today about is algorithmic bias and how algorithmic bias is exacerbating disparities worldwide, including access to truth, systematic exclusion, and bias. I also will discuss ways to overcome them.

Our world is being hacked. The scariest threats that are facing us are no longer overt threats. Malware distributed denial of service attacks, ransomware, doxing, et cetera, are still issues, but not our biggest problem. Our biggest problem is the subtle kind of hacking that has been going on. It is perfectly legal. It is well within the bounds of the systems that we engineers are creating. It is an entirely new kind of hacking that we never would have even envisioned even two or five years ago, but it is happening.

What does this hacking look like? One thing that has been going on that I am sure a lot of you are aware of is that ne'er-do-wellers have been systematically squelching truth with hoaxes and very subtle manipulations of Google page rank algorithms over the past decade. Many of these are state-sponsored actors. Many are engaging in a misinformation/disinformation campaign. This is also happening on social media outlets as well.

You probably heard about a lot of the political examples. There is a really great article in The Guardian recently about this, showing things like Holocaust denial and things like this, but there are a lot of other things you might not be aware of as well. For example, for vaccinations - which we know are safe and the safety is well-validated - the third hit on Google is something completely false. This is true for a lot of other types of health data as well - autism, Zika, and drug users.

We have a massive problem in how we are even

thinking about people from non-mainstream groups. This is a really big problem. It is only going to continue to get worse as we move forward.

What is the scope of this problem and why do we care? Well, if you think about it, back in 1973, this is actually a map of ARPANET. A map of the entire internet in the United States could fit on a single piece of paper. Interestingly, this is not where most people went to for their information. They went to physical libraries, where there were individuals who were stewards of information. They, of course, had their biases as well, but there was a notion of veracity and there was a notion of sort of some grounding of information in reality.

Unfortunately, what has been going on, just to demonstrate the scale of this to you, if you imagine sort of that ARPANET map, this is actually an image that is created by Jonathan Albright, who has been actually mapping out a lot of the systemic bias that has been going on. That is kind of - if you imagine that is one little tiny point, the scope and the scale is really vast. I think most people have difficulty even conceiving of the level of complexity that we are dealing with, but we really have to. 81 percent of Americans get their news and their information from websites and apps and social networking

sites. 90 percent of people get their health information from Google. This is quite bad.

It is even worse in developing countries. There is something called Zero Internet or Zero Rating, where people are provided with a free mobile phone and free access to data. It turns out that everything that people are seeing is actually being totally chuted by the ISP. For example, some people are only seeing Facebook. When people are asked about this, they believe that Facebook is the Internet.

There is another issue, too, though, in addition to this, which is that there is really a grave misconception even among computer scientists and roboticists that algorithms are inherently neutral and trustworthy. But they aren't. If you haven't read this book, I highly recommend it. It is called Weapons of Math Destruction. It is by Cathy O'Neil, who is a mathematician and a data scientist. It really outlines much of the darker sides of big data. There are a lot of really, really telling examples of how that data is being abused, being used to disproportionately discriminate against people from lower incomes, people from underrepresented groups, older adults, and so on.

One of the inherent problems in this - which

Missy Cummings has done a lot of great work on - is that people tend to over-trust autonomous systems. They tend to believe that somehow this computer is trustworthy. The worst part is that even our own technologists are also maintaining these beliefs.

There are a lot of examples. There have been cases of automated employment systems, which are systematically discriminating against people who complete non-evidence-based, completely biased health questionnaires. A lot of people say, of course, well, the algorithm told us so. It is not our fault we didn't hire all of these people with disabilities and so on. We also see these issues in cases where employers are creating optin or pay-up type of systems, which are again based on nonevidence-based data and, again, are poorly audited and lack transparency.

I think what we have seen a lot from sort of some of the great work that the Cathy O'Neils and others have done is that we have seen the scope of this with purely textual data. I am a roboticist. I deal with all kinds of sensor data, and all sorts of people in the world. Bias in these circumstances can be particularly troubling.

There are a lot of cases of facial recognition software being biased against non-Caucasian faces. There

was a recent article in the news about an individual who applied for a passport in New Zealand and was denied because he didn't "open his eyes". It was an automated system that was trained on Caucasian faces so it didn't recognize some of the characteristics of Asian faces.

There is another issue with gesture recognition. I actually was just talking with a company that was building a stroke rehabilitation system. They assumed that all of the users using this system were able-bodied and had their limbs intact and could fully move their arms and legs. This is a rehab system for stroke, which frequently presents with limb paralysis!

We have a lot of really major issues here. You have to imagine it is not just how we interact with these systems. It is also how these systems are using these data for things like automated housing denial, predictive policing, and so on. There are a lot of really bad dystopian futures, but unless we do something this problem is only going to get worse.

So what can we do? There are a couple of things we can do to help address this. The first is education. We need to educate our colleagues, our students, and our friends about algorithmic bias and about data bias. A lot of computer scientists I have talked to had no idea of the

extent that this bias was happening. This can happen through both formal and informal education. I started giving my students Weapons of Math Destruction. I am trying to organize some symposia on these topics. I love Missy's idea of bringing together technologists and policy individuals. That is another great way forward. This education needs to happen on all levels in all of our society.

The second thing we can do is to start Red Teaming for bias. In computer security, we frequently employ Red Teaming. We build technology and we check for points of intrusion. We check for data leakage. We check for vulnerabilities in our software. But we also need to Red Team for bias. This is not only bias within the system with perfect data and perfect users and perfect everything, but what happens when we have an onslaught of ne'er-dowellers? How does our system cope? What are the countermeasures to avoid these problems? How do we deal with bogus and biased sensor data? This is something that we really need to start thinking about.

The third solution is to keep humans in the loop. A lot of organizations believe that humans create more bias and they are not as good as automated technologies. This is not necessarily the case. Our data is flawed. Our

algorithms are flawed. Our AI is very, very dumb. We really need to get humans in the loop to check - to factcheck and verify information as much as possible.

Finally, we really need to require organizations to provide both algorithmic and data transparency. California has a new law that is requiring systems to disclose how public data are being stored and used. This is a fantastic start. I would like to see this expand a bit more. This would be really helpful to have this additional insight and transparency.

With that, I will close, since I think I only have about a minute left. I would like to end with this quote. Remember "Sneakers" from 1992? We really have to start thinking about our information and about data veracity as a critical important issue going forward. Thank you.

(Applause)

MS. MULLIGAN: I feel like I have been really well set up because Missy talked about having - being not allowed to take off until her hands were off. I am going to be talking about handoffs.





UC Berkeley School of Information

Handoffs with Integrity: Functional Fidelity and Values Protection in AI and Robotics

Deirdre K. Mulligan

Robotics and Artificial Intelligence: Policy Implications for the Next Decade December 12, 2016

Presentation of Deirdre Mulligan, 12/12/2016

What we are talking about today when we think about machine learning, AI, and robotics, is handing off the functions, sensing, sense-making, decision-making, acting, moving about in the world from humans to machines. We are thinking about complicated networks of these things. I am going to focus, in particular, on functional fidelity, so actually making sure that the thing that we hand off is the same thing that we thought we were handing off, that it functions the same way and some of the value implications of those handoffs. Then, in particular, focusing on some of the governance structures that need to be in place for us to do that. I am going to start with an old story and a new story, but the old story really came to life in this particular election. I was part of a NSF center that was studying transition from more paper-based to electronic voting systems many years ago. One of the things that we ended up doing with California, we did a top to bottom review of all of the voting systems in California. We found things like test mode code in them. That should make you a little scared.

Functional Fidelity





Presentation of Deirdre Mulligan, 12/12/2016

The thing I want to really highlight is that in transferring from paper-based to a direct record electronic voting system, one of the things that happened was that

there was an under-specification of the fact that the casting and the counting had to be discrete processes. What happened is in the first design of direct record electronic voting systems, you would go in. One, the ballot is rendered on the fly. It is not as though somebody has seen it and it is a fixed artifact. It has been rendered on the fly. You touch the screen. You look and you see, yes, that is what I want. You push the button. What do you think happens on the back end? Does it capture a whole image of that ballot? What do you think it did? It clicked. It incremented on.

Now, you think there has been a problem with the machine because 200,000 people came in and there are only 12 votes captured. You say we need a recount. What do you do? Nothing. You have literally, by transferring the function from paper to an electronic system without understanding that you needed software independence between the casting and the counting, you have done away with your possibility of actually ensuring the integrity of an election.

Now, why did that happen? Was that the engineer's fault? A lot of lawyers like to think that the engineers, they got it all - no. There is a lack of ability to talk and no one had clearly enough specified the

properties that we needed to transfer into a ballot. What is a ballot?

Flash forward. This is a Tesla S, which we all know. My mother will never get in one because somehow or other, it didn't see a truck. She says how could it not see a truck? Again, when you talk to people and you say the car sees, their understanding of the function of sight, how we see, and how the vision processing algorithms that are sucking in all the sensor data that is not just about vision - it is about radar and all different sorts of things - how they see - if humans don't understand the limits of the model, they may not trust the system. They thought it saw like they did and would see a big truck in front of it.

I want to talk a little bit and tease out some of the values conversations that both Missy and Laurel have already started.

Values: system



Presentation of Deirdre Mulligan, 12/12/2016

We can think about intentional bias. This could be a whole car presentation. This is a Volkswagen. They intentionally put test mode code into the car, undermining our ability to engage in regulatory oversight. That is one form of bias, intentional bias.

Two, there could be bias that creeps in through the process and values of the designers. The Jeep Cherokee, the one that Chris Valasek and his colleague took over remotely and were able to control braking and acceleration - not what you want to see in a car because, you know what, they all fail the same way, that particular model. You could remotely take over a whole fleet of heavy moving objects on the roadway. Not a good outcome. Cybersecurity, security, that was not a core component of the designers' mentality when they were developing cars. They are thinking about safety. They weren't necessarily thinking about the network as a point of vulnerability for the safety of the humans inside those cars.

Finally, Laurel also identified complexity. The Mirai botnet that we saw recently, hundreds of thousands of tiny, little, literally throw away objects invested with computing and computation power that were able to be leveraged in a more centralized way to attack in meaningful ways, our network infrastructure. Now, it is a very complex system in that we have all of these devices and people didn't understand the way in which they could be leveraged and networked to do relatively great harm.

Finally, in all three of these systems, we have privacy issues that are really changing in both their scope and also in more qualitative ways. First, these are all massive data collection systems. That is what they are. The way in which we are building our machine-learning algorithms is churning them over lots of test data, which is data about humans and their behavior and more and more their expressions and things.

Two, we see an increased interest in inferences. It is not just the information collected. It is what we can infer from it. Why that becomes such a complicated issue for privacy is that our legal models generally think that individuals are able to control what others know about them by exercising control over the data they disclose. Ιf we realize that through machine learning algorithms and AI, you can learn things about me that are quite distinct from what I think the information that I have provided you reveals, my ability to exercise control is out the window. Add to that the fact that many of the inferences that you are drawing about me are actually based on my mother's data, who I could not get not to send it in to 23andMe or my neighbor's data or my friends on Facebook. So this illusion of control.

Finally, AI is breaking barriers of place and skin that we have normally thought about as protecting privacy. All of these devices are coming into our homes, into our cars, places that historically, people have felt like are private places, where they are not subjected to massive data collection by third parties external to the home. More and more, people have had this idea - we are actually breaking barriers around people's thoughts, people's emotions, through data collection.

I want to just do a really quick - Laurel talked about a bunch of kind of bias in data.



Presentation of Deirdre Mulligan, 12/12/2016

These can come from lots of different things. It could be poor selection. It could be inaccuracies in the data. It could be selection bias, the population you are looking at. It could be bias in the patterns of collection. If we look at things like the predictive policing, the patterns of policing already reflect historic biases in who we have policed. All of these things can end up infecting our systems.

I am going to skip over and talk a little bit

Governance Common challenges • Transparency • Accountability • Bias • Privacy Mew or increased challenges posed by ML/AI/Robotics • Complexity • Dynamic systems—learning from data and experience • Personalization—what does it mean to be fair • Collapse of place and skin as structural protections

Presentation of Deirdre Mulligan, 12/12/2016

BERKELEY CENTER FOR

-X

So there are some common challenges that we have had with ICT traditionally: transparency, accountability, bias, and privacy. Some of these are new. There are some new and some increased challenges. Complexity/dynamic systems, like learning from data and experience how do we actually understand how they work. Personalization - what does it mean to be fair? What does it mean to not be biased in ways that are inappropriate when all we are about is personalization? This collapse of place and skin as structural protections.

Governance responses (taken and proposed)

- Improved standards, certification, testing (red team; accessibility; real world conditions)
- Transparency (code reviews; documentation reviews)
- Software independence (auditing machines w/ machines)
- Expertise in agencies and ngo sector
- Updatability of systems
- Participation—transparent and inclusive public process
- Security and privacy foregrounded



Presentation of Deirdre Mulligan, 12/12/2016

These are some of the responses that I have seen historically, over the years, whether it is to voting - you will see some of these are things that Laurel has mentioned. Red Team testing, accessibility testing, those were things that we saw emerge in the context of voting. Lots of calls for transparency, whether these are code reviews, document reviews, software independence.

Retooling Governance

 Communication and collaboration among lawyers, other domain experts, and engineers

 Technical standards integral to governance

 Public goods—privacy, cybersecurity

 market failures, illusion of control

 Regulators and public participation requires new tools and expertise

 Explainability (processes, rules, data)
 Experiential knowledge (simulations)
 Exposing assumptions and limits of models
 Public sector and ngo technologists

Presentation of Deirdre Mulligan, 12/12/2016

There is some retooling of governance structures that needs to happen. One, communication and collaboration among lawyers, other domain experts, and engineers is of utmost importance. This means that both technical standards are going to be more important, but so are things like explainability and interpretable systems.

We need to more consistently understand that privacy and cybersecurity are not individual interests that can be protected in an optimal way through market actions. These are public goods. They are going to be systemically under-produced. If you don't believe that now, I don't really know what else I can say.

Regulators and public participation is going to require new tools and expertise. I have mentioned now explainability, as have several others. Explainability is not just about being able to explain the rules, although I just came from the NIPS Conference, which is the largest machine learning and cognitive computing conference, and I was really heartened to see the sensitivity with which some of the speakers were talking about the need to actually not use the optimal model for getting the optimized results, but to use one that might be a little bit less optimal, but actually allowed them to reveal the rules to domain experts so that they could cross things out that either didn't make sense because of the domain experts' understanding of the data or were perhaps going to go in a direction that was not appropriate.

Experiential knowledge - people don't understand algorithms in an abstract way. We actually need to figure out ways to build simulations, other sorts of experiential learning that help people, regulators, but actual real people understand both the assumptions and the limits of the models. If we don't do that, it is a very bad outcome.

Finally, this was already said earlier today, too. The government needs computer scientists. Some of that is about hiring. Some of it, if you look at the autonomous vehicle policy that came out of the White House and NHTSA, they are looking to develop a kind of bench of experts that can advise them, that don't have to leave their day jobs. These can be academics and others. But it is not just the public sector. Civil society, in all of its permutation is also desperately in need of technical assistance. I think there is a real role at this point in time for engineering and computer science to think of the sort of interventions that happened in law schools in the 60's, where you actually create a career path for people who want to do science in the public good and that you figure out ways to support them. Thank you.

(Applause)

MR. GREELY: We have about 25 minutes for questions. I promised that there were lots of issues here. You may have been surprised that we didn't have four speakers talking about our upcoming robot overlords. The issues, the ways in which artificial intelligence and robotics are going to affect society, the ethical ways, the legal ways, the social changes, are vast. We could only scratch the surface, but we have 25 minutes for questioning to try to go deeper.

AUDIENCE: I wonder if the last speaker's last
chart didn't really have a manifesto of what the tech and ethics community need to work on. Those four points of explainability and so on are things that we could use to assess something.

MS. MULLIGAN: Used to Assess. So, I would be happy if there is a Mulligan Manifesto. That would be I do think that - I look at things like DARPA's fine. relatively recent program on explainable AI. I didn't know about that when I was talking about these, but I have since learned about it. I think that this issue of explainability and interpretability is one that is as machine learning is really hitting all different areas in the social sciences and the biosciences in deeper and deeper ways, people are really understanding that correlation is not sufficient. We need more theory. The minute you start building out models that have some more theory in them, you start to actually see things that are closer to explainable or interpretable. We need those in order to do some of the deep domain-specific work and get the kind of advances that we want.

I do think right now, there is much more of an emphasis on the interpretability of models. At NIPS just on Friday or Saturday, I don't know, there was a really great presentation talking about making the processes into data. You can understand the algorithm, what algorithm was chosen and what models, but if you don't understand all of the cleaning that happened to the data in the front-hand, you are still not going to be able to figure it out.

There is some nice alignment between collaboration across disciplines, reproducible research, which is obviously a huge push from this community, and the desire to make sure that these systems operate in a way that is kind of fair and promotes the values that we care about as a society. I am hopeful because the computer science community, the machine learning community seem to be looking for technical approaches to contribute to solving these problems, not just leaving them for people like me.

MS. FELDMAN: I think this issue goes back to a question the last panel was discussing, which is, "what is the role of humans in the future." I agree that a large portion of that role is going to be the one-on-one human interaction, the massage therapist they were talking about. However, I also wouldn't underestimate the higher skills that are needed for intermediation between the artificial intelligence world we may have and the human beings.

I was asked by a press outlet what my favorite app is. I said my son, Eli, because an app is only as good

as the techie who can make it do what I need. When I told that story at a faculty meeting, all of my colleagues came up and said I need an Eli. Those people who will help us make the technology work for us and do what we want will also be important.

MR. CALO: Something that hasn't come up much today and I think an ethical and legal panel will be a good place to raise it is something that is unique about robotics and artificial intelligence is the extent to which they, themselves, feel like people to us. The examples I gave this morning involving whether or not a toy represents something animate or - and there are many, many more - have to do with the way that this is a technology that uniquely, we are hardwired to interpret as being like a person. Ιt might raise questions about like if you are in a space that was historically private with something that feels to much like a social entity, are you really alone? Should we be leaving people with these social entities because it is better than leaving them alone, say in elder care? Are there ways that people can be manipulated because AI has many of the same social affordances as an individual, but also the capacity for infinite memory and persistence and so forth? I am thinking about work of like BJ Fogg at Stanford. Anyway, I wanted to get you to react to that

social dimension. Doesn't it feel like people and doesn't that matter?

DR. CUMMINGS: I would like to address that. There is a lot of work going on in social robots, affective robots. There have been the issues of whether or not you can turn off the robot and if the robot is begging for its life, will you turn it off? I would encourage you all to go look at that.

I am actually going to do a spin on that, Ryan, and say, actually, the thing that I am concerned about most, particularly as we go down in the drones/driverless cars world are the reverse social interactions, the negative social interactions that I think industries are woefully unprepared. I am a person who studies human interaction with technology and a person from North Carolina - I am originally from Tennessee, where we all learned to shoot at a very young age. My family is the first family that is going to let a drone, an Amazon drone, bring it a case of beer, drop the beer off, and then shoot it as soon as it leaves.

We laugh about it. We say, okay, so there is this whole other world. Are we worried about the anthropomorphizing, where we start to feel and have feelings for the robot. I think that is interesting. I think there are manipulations. In a very practical sense, a driverless car that goes into a bad part of some inner city with no human in the car is going to be robbed/attacked. This is just the reality. Google has seen what people will do to driverless cars in Palo Alto. Even though it might be prankster type of teen intent, there is this whole other problem of what humans do when they think that they are alone with technology.

I think it cuts both ways. I would actually tell you the social affective research community is focusing far more on the positive than the potentially manipulative influences. What happens if a driverless car goes into downtown Las Vegas from the hours of 12am to 4am, what kind of crazy things are going to happen because humans are not that reliable?

MS. MULLIGAN: I think there are definitely some challenges as robots get integrated into people's life. I would like to push back on the assumption that they are going to be freewheeling, moving around, physical robots. Robots can mean lots of different things.

How many of you have seen the movie WALL-E? For those of you who haven't, it is in the future. We basically have made this planet uninhabitable. We are all on this spaceship. We are moving around in these like

leisure chairs being fed way too much. It is not a pretty picture they paint.

Everybody gets very focused on the fact that these robots may be subversive. They may be trying to take us over and do bad things. At the end of the day, the thing that is really subversive that has actually been trying to make sure the people don't ever go back to planet earth because it is interested in its own well-being is the ship. They don't even perceive the ship as a robot because it is the built environment, but it has the same kind of AI and cognitive computing built into it.

Robots are going to take many different forms and will be able to nudge us and interact with us in lots of different ways. In many instances, the ones that I think historically have been most pernicious are those that recede into the background, that we don't even realize they are there. We don't know we can turn it off. We don't appreciate the fact that they are collecting the data or that they are altering our environment. The things that just kind of go into the background, I think, are the ones that we should be even more concerned about.

MS. FELDMAN: I love the way you framed your question in terms of what we think this item is and how we relate to it. The law moves so slowly and it moves by

analogy. Much of what we have is based on human players. As we move to thinking about how to regulate and what the legal systems are about that, we really need to shake that imagery. If we just borrow analogies about human beings in the law and apply them to artificial intelligent devices, we will go down some very strange paths.

DR. RIEK: That is a really good question. There is the issue of anthropomorphism, but there is also the issue of trust. There is a huge amount of research showing that people trust electronic systems more than humans, in terms of disclosure, in terms of what people reveal to the robot and things like that.

I worry a lot about - I work with people with cognitive impairments and older adults. I am really concerned about not just manipulation by the robots, but also by ne'er-do-wellers, who might be able to hack these things and alter them. We have talked a little bit about this, but it could be a major issue that could affect a lot of people.

DR. POLSKI: Margaret Polski, Krasnow Institute. Basically, if we are going to talk about regulating or governing or creating rules and laws, we have to talk about what it is we are trying to govern. This is, it seems to me, heterogeneous systems. That is what we don't know how

to govern, in part because we don't understand the behavior of heterogeneous systems and they are evolving. Where is the research agenda here? What do we need to start investigating and how do we need to start investigating it?

MS. MULLIGAN: At least one question - you speak from what you know. I did a qualitative research project where I talked to the people who were responsible for privacy and data protection in large companies in five different countries, Germany, Spain, U.S., France, and the UK. Part of the goal was to figure out like what they thought privacy meant and then how they operationalized it and to look at how that was produced by both kind of internal structural choices as well as external forces, such as formal law, the behavior of regulatory agencies, penalties, civil society, unions, whatever.

One of the interesting findings was that there was some value to having more ambiguous rules, as long as there was meaningful oversight and enforcement because it required firms to invest in professionals within the firm to help them sort out a constantly evolving, ambiguous external environment of risk for them. In doing so, it helped privacy weave its way deeper into the organization because it wasn't something that was just easily dealt with as legal process. It required engineering resources. It required buy-in from various business units. It became something that could actually be subject to design and reflection in products, not just papers and notices to consumers.

There is at least some, I think, qualitative work required to try to figure out - like if we want people to think about bias in the design of their machine learning algorithms, what sort of regulatory structures, what sorts of educational interventions, what sorts of incentives of other forms might best do that?

MS. FELDMAN: I think the research agenda has to be interdisciplinary. The law is used to thinking about dangers as coming from individual points. If you control that point, you control the danger. That is not an applicable analogy for the types of things you are talking about with understanding heterogeneous systems. Law needs to understand and borrow from systems theory better to understand what is happening. At the same time, you couldn't have a systems theorist come in and impose what would work in a legal system. You have to look in both directions.

DR. CUMMINGS: I would like to add a very specific one. That is academia, the research community, in general, is woefully deficient in testing and certification of

stochastic systems, period. The FAA would never let a plane with the new, automated landing capability ever carry passengers unless it went through a standardized certification process, which is in place today. I am not necessarily a huge fan of the FAA's policies, but at least they have a set of policies that have clear standard criteria that they apply.

This community is desperate for such standards across all domains of autonomy, including healthcare. The government regulatory bodies simply are ill-prepared - they cannot do effectively regulate it. In the case of NHTSA, NHTSA is pushing it down to the state level. NHTSA, through this vehicle policy that they just released, basically has said that all 50 states have to develop their own testing and certification profiles for driverless cars. If the federal government can't hire anybody in this field, I can guarantee you the 50 states cannot either.

MS. MULLIGAN: I would add one more thing. Again, I am going to - well, the idea that this is going to be a kind of complicated composition of humans and machines brings me to my first slide, right, about this issue of handoff. We need models that help us better understand how to optimally handoff for things like accountability. Part of what you are getting at here is we have this complex

system and things - in structuring our handoffs, we can actually make sure that accountability and responsibility, which doesn't always mean liability, but those things are transferred in ways that are meaningful so that we maintain the same level of understanding of how the system function.

To go back to my voting example, as a secretary of state, you are responsible for being able to say this election was fair and free and accurate. If all of a sudden, you are procuring machines that include contracts that prohibit you from doing meaningful testing - you can do logic and accuracy and nothing else, that preclude you from doing source code review or Red Team testing, you completely lost your ability to know whether or not you can still make a statement about that election. That is the place where people were when they were buying machines. It didn't have to be that way.

In part, I think Missy's statement about who we task with procuring systems or regulating them really matters. Again, there it was at the county level that we are making decisions about election technology. If you want to talk about under-resourced places, I don't think I can find you a better one.

MR. GREELY: Missy, if I could just follow up with you for a second and your focus on testing and

certification of stochastic systems, does that imply some kind of premarket approval regulatory scheme as is true with commercial airlines or the FDA and drugs?

DR. CUMMINGS: I think that is a great question because it is not clear, given the fact that machine learning's stochastic reasoning, probabilistic reasoning, and side systems can show up in a number of different places. I certainly think that we should, as a rule, if there is going to be a major change to any transportation system, which controls its three axes, if it is in the air, for example, if there is going to be a major change, then that system should have to go through some kind of certification process.

Now, the way it works in the FAA is if you do a major system change, then it is a major certification process. If you do a minor system change, then it is a minor process that doesn't have to meet the same rigor. I am not saying that NHTSA and/or the Federal Railroad Administration, if they ever get this far, should go exactly with the model that the FAA works on now. What I am saying is that there is a process in place and that we should be having that discussion right now at the federal level before you try to put it into the state's hands who are completely incapable of getting their arms around this

problem.

MS. MULLIGAN: Can I add one thing? In the voting area, there were standards and certification processes. What really matters is what is in those standards and how transparent are the certification processes, who is running them, et cetera. It is not just a question of whether we need them. There is a lot of devil in those details.

MR. GREELY: We have time for one or two more questions, maybe only time for one more answer.

MR. NEW: My name is Joshua New. I am with the Information Technology and Innovation Foundation's Center for Data Innovation. This is primarily towards Laurel. You had mentioned the idea of algorithm transparency. The need to prevent against all of these biases or undesirable outcomes, we need to make sure these are open and auditable. I understand the concern. I am completely with you on quarding against potential undesirable outcomes, but I think the idea of algorithm transparency has become like really popular. It is almost like a buzzword now. I think it is incredibly counterproductive towards the development of AI as a whole. Granted there is a lot of - I quess there are a couple of different definitions of it, but the most extreme, which is what I believe you were referencing, is the algorithmic systems need to make their data and

underlying algorithms available to the public so anyone can see them. Is that not the case?

DR. RIEK: I wasn't going quite that far. Here is a possible example. Let's say I am a company and I am building a piece of software and I am going to release it to the world. I bring in an outside panel of individuals who really Red Team the heck out of my code. They examine my code. They sign NDAs so we can still do intellectual property, but they actually go through a very rigorous testing process in a way that we are satisfied with, that we believe is providing us with a real good test of this system.

That is really what I think a lot of us are talking about. We have absolutely no checks and balances at all right now. There is zero. There is nothing. There is nothing preventing any company or any organization from using these systems even in ways that they don't intentionally do, but that end up getting usurped in ways they don't expect or predict. Part of the problem is that there is such a rush to push things out before taking the time to really do some of this examination and testing. I don't think it necessarily means all code should be open source, all data should be open source. I am not saying that. I am saying that there has to be a degree of

transparency that is provided in a way that we trust.

In general - not always, but in general - if we take a drug that has been tested by the FDA, we typically trust that it is safe. When I fill out a job application at McDonald's to go work there, do I trust that the application I am filling out is using good practices of employment? Do I trust that the system that is driving down the road is using really good pedestrian detection that is detecting people from multiple types of backgrounds and races, and detecting white trucks against white skies? We don't have any way to provide this trust - we have no confidence right now in any of these systems that we are employing.

MS. MULLIGAN: Nobody on this panel is saying we need a regulatory agency focused on AI. I think looking at as we move AI and robotics into different parts of the economy or different government activities, there are regulatory regimes that are in place. The question is how do we retool our governance structures to ensure that transparency, accountability, oversight over issues of all sorts of bias are maintained? It would be a very bad outcome if we are in the place where we were with the first round of direct record electronic voting machines.

MR. GREELY: Let's move on to one last question.

DR. BANAVAR: I wanted to kind of bring up the topic of professional ethics, which is a whole domain that I am worried about a lot because we go into so many industries and talk to different kinds of professionals, who have all established their own standards, sort of the standard Hippocratic oath kind of thing for healthcare and there are others and so forth. How do you see professional ethics being incorporated into decision support systems in AI, if you have thought about it? If not, do you feel like there is a kind of an approach that the ethics community has followed in human practice of these professions that we can learn from to decide how to do it for AI systems as well?

DR. RIEK: There are two things that have happened recently that are really exciting. One is the Communications of the ACM [Association for Computing Machinery] just updated its ethics policy for the first time in 25 years, which is really great. That is the professional organization for computer scientists. The second thing that has happened is IEEE [Institute of Electrical and Electronics Engineers] has a global initiative to look at ethics within autonomous systems. There are well over 150 people worldwide thinking about this and working on this. They are going to release the

document actually within the next couple of weeks. This is, of course, a work in progress, but we are actually specifically talking about professional ethics as a community and what we, as a community, can start to do, which I think is a good starting point.

MS. MULLIGAN: As I understood your question, it was in part like as a lawyer, if I am going to rely on a decision support system, how do we make sure that my ethical responsibilities are reflected in the system? I have mentioned a few times kind of making clear the assumptions and the limits of the models. Those are really important things.

I think one of the things that you find is people will over-rely on a system - oh, well, that is what came out of the machine. It must be right. I shouldn't override it. If they understand on the frontend what are the assumptions that are going in and what are some of the limits - so, take all of the autonomous vehicles. At some point, they say, oh, can't be in autonomy mode; throw it back to the driver. Well, why? What was the condition that failed that said the machine could no longer operate independently? I, as a driver, just learned something about the machine.

The 404 error code is not really useful. The

proliferation of error codes - right, there is a Fahrenheit 451 error code now. It means it has been taken down for legal reasons. Those sorts of error codes so that - it is a system. It is an algorithmic system. It is not just the algorithm. The more we can help domain professionals understand how the thing works and that improves their ability to make sure that they can continue to use that thing and behave ethically. It requires consultation on the frontend. It means you have interdisciplinary teams building things out, making decisions about data, making decisions about data cleaning, making decisions about appropriate algorithms. That is a lot of interaction. When you get out to the user population, assumptions and limits.

DR. RIEK: Another issue along sort of those lines is doing stakeholder design from the very beginning of the project. So, ethical-centered design, value-centered design - it is not like you start a whole project, you build this whole system, and then you worry about security and humans. You start at the very beginning. You include them in the process. I think that enables you to have a very good, robust system, which is going to help to be defensive against a lot of these sorts of issues.

MR. GREELY: We could go on for hours and hours,

except that our stomachs may not be willing to do that. Please join me in thanking our great panel.

(Applause)

(Luncheon recess.)

AFTERNOON SESSION (1:35 P.M.)

Agenda Item: Panel 3: Security Implications: implications for public safety and national security

MR. CALO: Hello. Welcome back. This particular panel is going to be focusing on the security implications of artificial intelligence and robotics. We're going to be changing up the format a little bit, and rather than having a series of presentations for this particular panel, which is a little shorter than the others, we're basically going to be engaged in a conversation.

So I'm going to start asking some questions that I have for our panelists, who I'll introduce in a moment. And then I'm going to open up at the soonest possible time to have you all jump in as well with your questions.

Again, this is going to be about safety, about basic notions about how to if not guarantee, then maximize safety in a variety of contexts in light of advancements in artificial intelligence and robotics.

We have three great panelists here to help us through that. I'm not going to introduce them at great length, as I might because their bios are available to you. But immediately to my right is Stuart Russell, who is at Berkeley, and actually has an entire new center that I hope you wind up telling us a little bit more about, that is focused on how to build AI that is safe and sustainable, and of course a longtime practitioner of artificial intelligence and the author of one of its foremost texts.

To his right is Gaurav Sukhatme. And he is at the University of Southern California, and also works on artificial intelligence and robotics, has a wide variety of experiences; but I understand that recently has been focusing on a lot of underwater robotics, and the like. But I hope that your work comes out as well.

And then to his right is Doug Maughan. He is Division Director for the Cybersecurity Division in the Homeland Security's Advanced Research Projects Agency, HSARPA. I actually spent some time, I was staying with DARPA on Friday, and you are the DARPA of Homeland Security. Domestic DARPA, I guess you would say. It doesn't seem like he agrees with that characterization, but I'm

delighted you could join us and I really, really appreciate it a lot.

Okay, so in many ways the previous panel and the one before it have really teed up this set of issues we're going about here. In particular, I want to start us where we left off with some of Missy's comments, Missy Cummings' comments, about just how hard it is, ultimately, to ensure that complex systems are safe.

One of the questions that I'll actually open up with Gaurav with this, if I may, goes directly to Missy's point. What is the consequence of having all of these systems -- here we are, pushing for them; we're asking for them to save us from train wrecks and the like, and we're thinking about more and more robotics and AI in our midst, and controlling cyber-physical systems -- how do you begin to validate that they're safe enough to be amongst us? Start there.

DR. SUKHATME: Thanks very much. I think one of the challenges in this space is that as these AI and robotic systems enter our midst and be used on a day to day basis, there is the potential not only for inadvertently leaking data or making information about ourselves available to people we don't want it to be shared with, but I think when it comes to robots in particular, physically

embodied things that you interact with that are in your midst, there is the increased prospect of malicious actors being able to actually subvert them for their own uses in ways that can be very destructive.

So I think there are tremendous security implications that we haven't thought through. When we think in great detail about safety implications for day to day objects and highly regulate it, these highly customizable, software-driven artifacts that we are bringing into our midst that people can subvert to their own uses. And this, I think, raises a lot of concerns.

A lot of toys come with software these days, and the things you can do with them are not the things they were originally intended for. So I think there's a whole opening up there of safety and security considerations.

MR. CALO: Let me push you on that, right? So lots of people have been hurt by lots of toys. I teach product liability law. We spend three days on it. These things are dangerous. They're not usually hurt because someone hacks into the toy; they're hurt because they tried to eat a ball or something. So I wonder -- are you implying or saying that the major risk is going to be a cybersecurity threat? Isn't the starting point, just making sure that these things are safe to do what they're supposed to do, even in the absence of interference? Do you have a sense of what the priorities should be?

DR. SUKHATME: I think you are right. I think the starting point is exactly to make sure they behave as advertised, and that they at a basic level do what they are supposed to be doing. Giving those assurances is nontrivial, for systems driven largely by software. I think the early panel talked about this. Even giving a condensed, small explanation of what a system is doing, if it has modern AI in it, it's relatively difficult to do.

I make this point sometimes when I teach roboticists is that to some extent, machines have to meet a higher bar in this regard. Somebody, I think, gave the example earlier on of the human Go player not being able to say how he does it. We have a convenient word for it; we say he's a genius. A machine doesn't have that ability, at least not -- it wouldn't make us happy if the machine said, Sorry. That's too difficult for you to understand. I can't explain it. Just a genius. That's not acceptable, at least I suspect, to most people in our generation. That's not an acceptable output.

In a way, that's a difficult problem. People who work in explainable systems try and address this in a deep way. It's a difficult thing in its own right.

I think you're right. At a basic level, you simply want your system to be able to say what it can do, and give a guarantee that it will do that. But I think there are some emerging concerns, that some of these things can be subverted. When I sometimes say things can be easily weaponized, it may seem like I'm being alarmist. But I do worry about them being taken over for purposes they were not designed for. And I think that's a concern. When you look at things like quadcopters. At dinner yesterday, Stuart made this point that quadcopters are things that are great toys, and they're entering the mass market. And you can do very bad things with them. Very, very bad things with them -- cheaply, efficiently, ruthlessly. And I think that's a maybe something a little further down the line, but it may not be as far away as we think.

MR. CALO: Well, you name-checked your neighbor Stuart, so let's bring him into this conversation. This is an immediate follow-up thought. I would like to get you to react, Stuart, to the following taxonomy/characterization, which is that we have to figure out how safe these things need to be before they're let into the general population, right?

Missy was talking about the FAA's procedures to certify planes, and so forth. Did you know that for

mission-critical components, for things that actually keep those planes up in the air, they have to be certified to 10 to the negative nine adverse event. That's probably not appropriate for a drone. It's probably not appropriate for a car. So first they're setting how safe they have to be before they enter our midst. And then there's various techniques that you would use in order to achieve that safety, to validate the software against some standard. It's explainable in a way that we can unpack and figure out in advance and diagnostically what went wrong.

And then there's the added fact that these are cyberphysical systems and so you also have to do the same sort of exercise with the adequacy of the security. I wonder if I can start with Stuart, and get you each to react to that. Is that model complete? Does that make sense?

DR. RUSSELL: I think we have to distinguish between what we might call local safety and systemic safety. So local safety would be failure of one selfdriving car or one surgical robot causing one or two or a small number of casualties. And there, I think it's a serious concern, obviously, for the people who are dead -and we saw this already with the Tesla, and Tesla reacted by pulling back on the level of autonomy that they would allow.

But very quickly, you realized that the product is having a higher rate of accidents than is acceptable, and it'll be taken off the market before it has very serious negative consequences. But the systemic risks are things where it could have a societal-scale impact.

As Gaurav mentioned, I think about the weaponization of AI, the creation of autonomous weapons as one of these systemic risks because there, if it's having lethal effects, then rather than pulling it off the market, we'll make more of them.

The thing about autonomy, by definition, is that when a weapon is autonomous, it doesn't need a human being to manage its activity, to supervise and execute each of the attacks that it carries out. So you get this decoupling effect. In computer science terms, you need order one people to carry out order n attacks. Essentially, for any n, you can have 10 million or a billion weapons being launched and managed by a single person or a small group of people.

So in that sense, it creates weapons of mass destruction. And to me, this is the main concern in the near term, for what a negative impacts on society might be for AI. I'm not too worried about should a car run over the

old lady or the little child in the pram, the trolley problems, those kind of things. Those will get sorted out. And if they don't get sorted out, then we won't have those products for a while. But these more systemic impacts, I think we have to take very seriously.

MR. CALO: One person might dive into an outdoor pool and hit their head, but this is like everybody diving at the same time and hitting their head. I take your point about the systemic. It's almost as if we should have agency, like a federal agency, that is devoted to protecting us to these kinds of threats.

Anyway, Douglas, I thought I could bring you in on this question of is this something that you guys are thinking about? This sounds awfully scary to an awful lot of Americans and others. And at the end of the day, is something that your agency thinks about, and how do you think about it?

DR. MAUGHAN: So we are the swimming pool checkers? Is that it?

MR. CALO: In my bad analogy, right.

DR. MAUGHAN: Actually, in your opening statement there, and even Stuart's comment, you talk about things that have AI or robotics, and how we think about them. And I think about it from the standpoint of at the core, it's

software. Software controlling systems, but if we think about where and even Gaurav and his comments, and things being attacked for malicious purposes, are we having the conversation in AI and robotics about the quality of software engineering that's going into these systems? Because in the end, we're not building secure systems today; are we all of a sudden going to build secure systems with AI and robotics?

From our standpoint, all of the systems you look at, whether it's cars, planes, dams, energy sector -- you name it; at the core, it's software driving some kind of system. That's one of the things we're thinking about is how do we make better software. Because in the end -- I don't care if there's anybody here that's a programmer -they still write buggy code. It doesn't matter.

How do we do a better job? I think the more recent thing they've done with DARPA at the High-Assurance Cyber Military Systems (HACMS) program, where they're looking at doing formal methods and faster formal methods, in order to validate and secure small bits of code -- it's a good start in the right direction.

MR. CALO: You know, Dennis, it occurs to me that a lot of folks in this room may not totally understand the scope of the work of DHS's ARPA, and having the

director here is a wonderful opportunity. Could you tell us a little bit about what your mission is and how it intersects?

DR. MAUGHAN: Sure. If you are not familiar with DHS, I'm sure most of you are pretty familiar, you see at the airport, and at the border. But you probably don't know what else is in DHS, right?

So DHS is the largest law enforcement agency in the government, with the Secret Service, CBP, ICE, and Coast Guard. We also have responsibility for critical infrastructure. So all of the sectors -- there are some sectors where there's a different agency like Treasury is responsible for banking and finance, but we still have the responsibility to ensure that 16 sectors as defined by the government are safe and secure.

We have the responsibility to secure government websites (.gov) -- so yes, we are responsible for the hack that happened on OPM. Even our own systems are vulnerable. We're responsible in FEMA. Every time there's a natural disaster, DHS jumps in and is responsible. The largest, most expensive natural disaster that happens every year? Floods. How do I use AI to predict the next flood so that I can make sure that people don't get caught in front of it, people lose less property/less money, etc.

So from a department perspective, we're 250,000 people, give or take, and have a pretty broad mission, not counting the ones you're used to from the airport and the border.

MR. CALO: And then what about your specific group?

DR. MAUGHAN: My specific group -- I'm the division director for cybersecurity within science and technology. So we're worried about creating technologies to help all of those operational guys with new technologies and more secure technologies.

MR. CALO: One question I have for you, and then I think a nice roadmap for this might be to talk a little bit more about the domestic challenges and more into, ultimately, the broader theater. But I thought I would start with -- what about actual equipment for law enforcement, both in terms of your own agents, and then also local law enforcement? Because as I understand it, DHS has some role in giving guidance in local law enforcement. Would it be appropriate to deploy some of these tools in those contexts? And what kind of safeguards would be needed if you did?

DR. MAUGHAN: Just like everybody else, the law enforcement community is dying from the volume of data that

they get. Criminals are online, in case you didn't know that. More and more of them are online. That's not as dangerous; they can make just as much or more money, and they're all doing it from their basement in their robe and slippers, right?

The biggest problems that law enforcement guys have is too much data, not enough technologies to help them do the data analysis. And then once you have that, then they can go do their traditional law enforcement job, which is knocking down doors and arresting people. But you've got to look at -- there's tremendous pattern analogy that you can do in all the data. Criminals are criminals; they don't usually change their behavior. They do it one way today, they'll do it the same way tomorrow. So there are some interesting things where we could take some of the computer science and the AI and those kinds of things in the background and apply them.

I think from a protection standpoint, I don't know if the law enforcement community has any special protections. The question is are we actually -- some of you have seen the National Privacy Research Strategy that came out of the White House this past year; some of those same privacy requirements apply to the law enforcement community. They have to protect our information. They have

to be US citizens, that type of stuff. Same thing applies. So as long as we're, in general, applying some of those privacy safeguards, then they apply to the law enforcement community.

MR. CALO: I have one follow-up question about that.

You talk about data, but there's also a kinetic, a conversation about how robots are being used in policing. I cannot tell you how many calls I got about using a police robot to kill somebody in Dallas. That was a huge deal. It's not quite about pattern recognition or anything else like that; it's just the fact about using a robot. Do you have guidance about how to use robots, and do you talk to local law enforcement about how they use robots? Is that a role that you could see?

DR. MAUGHAN: We have had certainly conversations about that problem, and I think it comes back to control. How do I guarantee that the robot -- *something* happened, right? What was the NTSB analysis of the robot, how it was used and what happened?

MR. CALO: As you recall, this was a purpose of use where they put a bomb on it and blew it up. I've gotten a lot of pushback from many quarters by arguing, Well, this is like a sniper, right? Sure, the robot can't be hurt by

the attacker, but nor could a sniper, and if the sniper were authorized, why can't we use a robot?

And the response was very visceral. Because it's a robot. You can't blow people up with robots. That's ridiculous, right?

At a minimum, I think we need to be thoughtful about when robots, even when they're tele-operated, would have some kind of kinetic effect and actually hurt.

DR. MAUGHAN: I think in the last panel, there was certainly some discussion there about some of the policy issues, right? So what are our policies as it applies to using these types of technologies, and especially in the policing world? What's more important -- making sure that my policeman is safe, or making sure, taking the robot and using it as a weapon-carrying instrument to take the person out?

That's an interesting policy debate that -- we haven't had those conversations, when you start talking about using robots. We've seen them in movies, but I don't think we've had them in real life conversations.

MR. CALO: Thank you very much. I'm glad I had a chance to ask that. I've been thinking about that for some time.

So Gaurav, I want to get to you and talk a little

bit about the evolution of the concept of security in general, in the sense of national security, and how it might evolve with robotics/artificial intelligence.

Typically, as you think of security as being predicated at least at one level, on a kind of having an asymmetric advantage, because it is a deterrent, and it also allows you to respond. So if there's a scenario that Stuart describes where all of a sudden, someone takes over all the cart -- but the other side of that might be, Can you really preserve asymmetric advantage in the theater of war in light of artificial intelligence and robotics? Or will it be even easier to do so?

DR. SUKHATME: Yes. I think this, preserving this asymmetry is going to be challenging. And I think it's going to be challenging for the reasons that Doug just mentioned, which is at some level -- I'm oversimplifying -but I think at some level all security is rapidly become cybersecurity.

I don't mean that in a trivial way. I think, obviously there are aspects of security that don't happen unless you have hardware. But the increasing degree in which the hardware that becomes important is under software control, and often these entities are networked, I think basically means that preserving this asymmetric advantage

is going to be harder and harder. If the software on your new weapon can be easily hacked, then in some sense, the weapon can be taken over much more easily?

The reach of people's ability to get into other people's software is dramatic, it's dramatically more. It's an action at a distance that your adversary has that they don't enjoy with other kinds of weapons. And that is worrisome. I think the implications for security, for national security, are very significant in this respect. I may not have a solution to it, but I worry about it. I think it changes the way we think about security, and the way you would go about preserving that advantage.

We need to think very deeply about how to address this.

MR. CALO: I think Stuart wants to jump in. I was going to just clarify -- do you mean literally, that we'll have this weapons advantage that is, in terms of superior firepower, more personnel and so forth, but that somebody, an adversary would actually be in a position to neutralize those weapons in a way that they really couldn't before without a similar investment in that kind of hardware.

Is that what you're getting at?

DR. SUKHATME: Yes, I think it would not potentially worse than neutralize, but to neutralize or to

subvert, and in fact employ those weapons against us. To some extent, the US practices these techniques to the extent that if you believe the reports about how the centrifuges in Iran were damaged by the introduction of a software hack were true, this way of engaging with the -you know, it costs a lot of money to build a centrifuge and to build a nuclear reactor. And it potentially may not cost that much money to embed two lines of code in it that causes it to spin at a rate that self-destructs.

So now, if you take this to weapons systems and presumably there are some fairly malicious consequences people can have at relatively low cost. And I think there is a danger there of weapons being subverted.

MR. CALO: Stuart, please. Sorry to interrupt.

DR. RUSSELL: There are at least two serious ramifications of this. So one is that in the good old days, you could buy the latest editions of Jane's Fighting Ships, or Jane's Armed Forces. These are enormous compendia detailing for each country how many tanks of a certain model they have, how many aircraft, how many ships, how many soldiers. And you had a pretty good estimate of the balance of power, regionally and globally.

We knew how many missiles the Russians had pointing at us, and we had pointing at them. But if your
armed forces are largely dependent on robotized autonomous weapons that could have been subverted -- in some sense, you don't even know how many weapons you have. We might think that we have, if they ever come off the production line, 740 F-35s. But we might not have any. We might think they're F-35s, but in fact they're not, because they won't do what an F-35 is supposed to do when we ask them to do it.

So that creates this huge uncertainty. And it might not even be that they're subverted in the sense of changing how they behave, like making them attack us for example. But simply knowing what algorithms they run -easy enough to render them much, much less effective because you can predict how they're going to behave, and that makes them extremely vulnerable.

So I think the degree of uncertainty that this introduces into whether or not there is a strategic balance or an asymmetric advantage has really serious consequences for how we think about the game theory of international relations and security.

MR. CALO: Just to push back on that a little. Why don't we think that maybe, for example, the United States, would just develop superior capabilities there as it has in the kinetic world? There's a story I read in The

Atlantic this month talking about -- this may be apocryphal -- but the idea was that the president of China came and had a conversation with President Obama. President Obama made some comment that we don't know what it is, but this corporate espionage has to stop, and according to this article, made a veiled threat, We're going to use our capabilities offensively potentially. And then supposedly the activity dropped off.

Why wouldn't we just expect that if we devote enough resources to it, we'll gain superiority in this universe too? Is there something endemic about this problem, or is it just a matter of, Gosh, we really haven't caught up. We haven't invested enough in it yet.

DR. RUSSELL: One of the reasons why people are talking about autonomous weapons as the third offset is precisely that we no longer have the advantage in kinetic weapons. So the second offset being precision-guided munitions, the ability to basically destroy anything that we could protect anywhere on the planet pretty much, that advantage has evaporated.

And when you look at what are the pieces of this third offset, the autonomous weapons, honestly it's not that complicated. It's easier, I think, to build an effective autonomous weapon than it is to build a safe

self-driving car. A safe self-driving car has to have maybe not nine-ninths of reliability like the aircraft, but eight-ninths of reliability. A very effective weapon could have one-ninth of reliability, and still be thoroughly above average for a typical military system.

It's pretty much out there in the public domain already, the technology you need to build an effective autonomous weapons system. Right now, if you look at quadcopters, I think the best quadcopters are probably Chinese. And then the rest is software. So where is our decade-long military advantage in this arena? I don't see it.

MR. CALO: Other comments?

DR. SUKHATME: I agree with Stuart. I think particularly has hardware levels off in some domains -it's not going to level off everywhere. But the moment it levels off, I don't see that we'll have asymmetric advantage in the software, particularly when the pieces of software used to weaponize some of these systems are in fact nearly commodity software. Nearly, or at least the infrastructure needed to build them is highly commoditized, and is really freely available to all at zero cost. I think the advantage is very, very negligible and there's a potential there for tremendous harm as a consequence. MR. CALO: Douglas, if I may, how do you prepare for a world like that? Assuming that this is even a viable scenario, we have several experts agreeing about the potential that some of these asymmetric advantage will dwindle, how do you prepare for a world like that?

DR. MAUGHAN: That is an interesting question. I do agree with Stuart and Gaurav in the sense that 25 years ago, when it was purely a kinetic world, we had the advantage, and we won. We proved that in Iraq. But in today's world, there are no longer only kinetic -- and the other problem we have is those who are our enemies can more easily buy capability that they wouldn't have had otherwise.

How do we prepare for it? Well, I certainly believe that there's an international piece to this. I think it was briefly mentioned in the last panel. But in the end, self-destruction of everyone is not a really good idea. But we have to continue to prepare ourselves to be able to do the offensive piece. From the Homeland Security perspective, our job is to try to worry about doing the defense of the homeland. And so you still have to have the autonomous capabilities to try to do defense. There's a separate set of things more primarily hinted to in the last two remarks, which is the autonomous ability to take the

offensive. But that's not our mission in Homeland Security. So now the question is, How do I defend my critical infrastructure? How do I defend our own cyberspace when in fact our cyberspace is what is the battlefield being used by our adversaries as well as ourselves?

MR. CALO: Stuart, just to put this to you directly, how feasible are fully autonomous weapons? I'm detecting that you're saying in a sense that some sets of them are really readily feasible, right? They're sort of almost trivial. But what about larger scale -- how feasible are we getting to a universe where a lot of these decisions are actually made autonomously? And further, what repercussion does that have?

DR. RUSSELL: I think there are already weapons that have crossed the line into autonomy. So when we talk about autonomy, we're talking about autonomy not in the sense that human beings think about autonomy -- I can choose to be whatever I want, a Dr. Seuss kind of autonomy; it's more like the autonomy that we grant to a chess program. We don't tell the chess program what to do; basically the objective of the program is to win, and invariably, within a few moves of the beginning of the game, we get to a situation that has never been seen before. And the program is entirely on its own, using its

own decision making to decide where to move the pieces, and which enemy pieces to capture. So that level of autonomy is all we need for weapons to be autonomous.

I think there's been a lot of obfuscation in the diplomatic circles, talking about autonomous weapons as well. They have to be self-aware. It could be decades and decades into the future. But there are already weapons that have the capabilities. For example, Israel has the Harop missile, where you simply specify a geographical region, you specify a target criterion, which could be visual, or based on radar frequencies or something like that, and the weapon will hunt around this geographical region until it finds something that satisfies the criteria and it will destroy it. So it could be a tank that it recognizes using its camera, or it could be anti-aircraft defense installation that it recognizes based on a radar signature.

And I believe in Azerbaijan, recently it was a bus full of people that it destroyed. The UK Ministry of Defence says that in uncluttered scenarios which includes aerial combat, naval undersea warfare, full autonomy is feasible now. So this is not science fiction, this is not decades away, and I think if we had a Manhattan Project, I think we could be fielding millions of weapons within 18 months to two years.

But as Missy mentioned earlier, one of the biggest problems is that DoD just doesn't have access to sufficiently talented technical people to do this. So it really would have to be a Manhattan Project, where you actually in some sense conscripted scientists, which might happen if there was a sufficiently serious national emergency as we obviously had in World War II.

I think the consequences are similar to the consequences that we avoided with biological weapons. Reading the history of biological weapons is very interesting. The US had a very large program -- in today's terms, it would be in the billions of dollars per year being spent to develop biological weapons. For example, people were trying to develop disease agents that would only kill certain types of human being, Slavic types being one target, and work on developing antidotes so that the all the US population would take the antidote, and then we could wipe out everybody else.

What was pointed out, initially by scientists and then Henry Kissinger took this on to convince Nixon, was that if you create this technology, it's a kind of weapon of mass destruction that would be much cheaper, much lowertech, than nuclear weapons; much easier for many nations to acquire and deploy. You would be reducing the national

security of the United States considerably by doing that. So Nixon agreed with this argument, and he unilaterally cancelled the biological weapons program. And then a couple of years later, there was the Biological Weapons convention. It's true that the Russians continued to cheat and worked on their biological weapons for another couple of decades. But we haven't had a war in which biological weapons have been used. So it's been fairly successful.

I think autonomous weapons have the same capacity. They can be low cost. So it's a simple calculation. You look at a quadcopter that's one inch in diameter. That can carry enough high explosive to kill a person quite easily. You can put three million of those into one container. It can go faster than a human being can run. And they have a range of a couple miles. So you can wipe out a middle-size city just as you can with a nuclear weapon. But you could also say, I'm just going to wipe out all the males between 12 and 60. Leave all the females, leave all the kids. So it becomes a weapon that is an extremely attractive, effective, cheap weapon of mass destruction that pretty much anyone, as long as they can afford a few million dollars -- I mean, these things might only be ten bucks each. As it turns out, it takes about \$1000 worth of lead to kill one human being in warfare;

somewhere around 3000 bullets were fired per human casualty in World War II. And you compare that -- if you have one of these, it costs \$10 and in half the time it finds the target and kills them, then it's about 100 times more efficient than bullets.

So it seems like you'd be creating a world where everyone would be worse off because these would proliferate and they would get used, just as Kalashnikovs have proliferated and they get used, and they are making the lives of millions of people miserable all over the world. This would be 1000 times works.

MR. CALO: I write in a lot of areas of robotics law, and I don't write in this area. And the reason is, because in the areas that I write with robotics law, I get to scare people, and then I get to say, But here are some solutions that we could do. But I feel like there are as not as many solutions here. You get really scared, and then people say, Well what do you do about it? And then there's no sense of what to do about it.

I want to push back in a couple different ways. First of all, it seems to me that if you could make really cheap autonomous weapons, or even really expensive autonomous weapons -- and I think there's a little bit of tension there, Stuart because on the one hand you are saying that DoD needs a Manhattan Project with all the best people in the world, but on the other hand you're saying, I could do it in my garage. I'm not sure where you fall on that. But we can come back to that in a moment.

But couldn't you just as effectively create defensive mechanisms? Do you see what I mean? So is one way to pursue this, isn't this just a standard way, and I'm trying to get at what's new here, is it the standard way of just, We would have just a bunch of nanobots that we're protecting our cities, so if their bots came, we would deploy our bots, and they would cancel each other out. The Gates Foundation gave a bunch of money for a laser that kills mosquitoes by hitting their wings, and then they don't spread malaria.

Is anyone on this panel, and I'll start with Stuart, confident that we can come up with countermeasures that are as effective as the measures that are deployed against us? And then if you wish to address the tension that I'm outlining, the idea that it's inexpensive, so inexpensive that we should worry about advancing it, but that it would take a lot of our national brainpower in order to do it.

DR. RUSSELL: I think the Manhattan Project would be how quickly could you get to a mass-manufactured, widely

deployed system. Whether that happens depends a lot -- it isn't just a matter of the technology advancing to the point where we can produce them. The way weapons systems work is it's not that companies decide, Oh, that's a really cool weapons system. I'll make some of those, and then maybe I can find some customers. They are usually commissioned by governments. And at the moment, governments -- at least in the Western countries -- have not commissioned those kinds of weapons.

In fact, some companies, such as BAE Systems, which is the second largest defense manufacturer, actually declared that they would never make such weapons. That was quite a positive development.

So one answer is that there should be a treaty. The question you raise is, What about defense? Lasers for shooting down mosquitoes is a great idea. If you could have lasers for shooting down microUAVs that would be also be a great idea. And that would not be anything that would be banned by treaty, because that's not a lethal weapons system. It's a defensive system that attacks robots, so that's completely fine.

We are spending a lot of money. The outcome of those research programs, which is as far as I know classified, so I can't say how much progress has been made.

They have been going on for 16 or 17 years. The so-called Black Dot program to develop anti-UAV weapons that are sufficiently effective to really give you blanket defense, but I don't know where they are.

MR. CALO: I want to eventually open it up to all of your questions and I have plenty of time. And I have one more for the group. But before I do, does anyone else want to react to that? To the defensive capabilities, to other scenarios you're concerned about?

DR. SUKHATME: I think countermeasures can be developed. And I think if you take them seriously, they certainly can be developed in some settings. But I think proliferation of some of these techniques may happen at a speed and at a scale that we haven't seen with other things. I think the lessons from biological weapons are instructive, as Stuart pointed out. It's probably the right to place to look for what can be done, because I think the threat from rapid proliferation here is high. Some of the key actors making the hardware that supports these are not in the US, and the software can be put together in many, many different parts of the world. I think there's a difference between the biological systems to look and see what can be done. The weapons are unconventional -- the supposed weapons are unconventional in that respect. MR. CALO: Thank you. Douglas?

DR. MAUGHAN: I would just add -- if you haven't read the book by Representative Michael McCaul called Failure of Imagination, his first scenario is a drone swarm attack on the Capitol. And this is what I think the key piece here is: we don't actually know how our adversaries might use large quantities of these. We think of them in how we might use them, and so we assume our adversary will as well.

I think everything that's been said is correct. There's work going on, certainly at the classified level. There are other things. We've been looking at it from the standpoint of how do I use these technologies to actually do good things? So how do I provide a drone in the desert for the Customs and Border Protection guy? He's out on the border; how does he see over the horizon, where there could be ten or 20 people coming over the border with guns, and he's by himself? How do I use that and that mechanism? So I think there are still good uses for the technology from our perspective, but I think there are plenty of scenarios that we have to think about how do we counter these unmanned systems that are in the hands of people who want to do us harm?

MR. CALO: Here is my last question, and then I'm

going to open it up to everybody. It's perfect because you brought up this book, Failure of Imagination, which I haven't read and I intend to. But as long as we're talking about imagination here for a moment, how worried are you about really catastrophic applications of artificial intelligence and robotics that truly gets out of our control?

People from Elon Musk to some of our greatest entrepreneurs and scientists, not necessarily all of them computer scientists, have invoked this idea of we're summoning a demon by developing artificial intelligence. Artificial intelligence will be our last invention is something that you hear, which is particularly foreboding, not because it takes all of our jobs like Martin worries about, but because it takes all of our lives for reasons that are not ... so anyway, as long as we're imagining for the moment, is this something that we really, really should be concerned about, and if so, what do we do about that?

And after your responses, I'll open it up to all of your questions. Do you want to start, Douglas?

DR. MAUGHAN: I am not a big movie watcher, but we've certainly seen all the Hollywood scenarios, whether it's Star Wars, or whatever, Attack of the Drones. But I think we're just still at the tip of the iceberg in how a lot of this autonomy and autonomous systems are going to be used potentially against us.

We've seen, as you said earlier, maybe Stuart said earlier, these cases of ones or twos where systems have gone awry. I think there's going to have to be, and it depends on the environment and the situation, there might need to be policy and regulation brought into some of these discussions. I don't think we're quite there yet, where we need to go running away screaming that we're all going to die. But I certainly believe that -- again, I come back to the core of artificial intelligence in the end, is being driven by humans who are writing software to behave in an autonomous fashion. So how do I worry about that from both a production of good things as well as production of bad things? And people are going to use things for good and bad, however they choose.

DR. SUKHATME: I think there is also a generational divide, as somebody in an earlier panel alluded. The degree to which one is comfortable in a built environment in which one lives has changed over time. I suspect someone who lived 2000 years ago, would be immensely uncomfortable with our way of life today. It would strike them as unhuman almost, because we are surrounded by built artifacts that we've created that we

all have come to know and love for the most part, and they would be utterly alien, completely foreign, and sort of shocking and scary to people who didn't know how they work, presumably.

Now there is this question about whether these new artifacts are different, or whether we will do what we've always done, which is to just adapt to our machines. We're very good at adapting to technology. So far, we've been very good at adapting to technology. But it has always generally been the case that we've told ourselves the story that the machines are not good at certain things that make us human.

And now, I think people are worried because there's some question about whether that's in fact true. And if the question is really bothering people, then maybe we will not be very comfortable with these technologies, and we will be summoning the demons.

But I don't know the answer at all. I'm not a philosopher and don't know this. But it is -- I don't know that people have said things like that. It's quite one thing to say that the invention of new machines will put people out of work, and people have said that repeatedly; one can take a position that well it's not the same this time around. It's quite another to say that these machines

will redefine what it means to be human. And people didn't say that necessarily quite so frequently in the past, and they say that more now. And maybe that is what makes it different. One ramification of it will be in the labor market, and a few other things, where there may be more profound ramifications.

MR. CALO: I have to say some of the computer scientists that I talk to, like Oren Etzioni, very serious -- he runs the Allen Institute for Artificial Intelligence. He just says no way. And Andrew Ng from Stanford -- well, formerly Stanford, now in industry as we talked about earlier about what's happening to everybody in the field -said something like we should worry about that like we should worry about overpopulation on Mars.

And Pedro Domingos (from the University of Washington) says it's not that artificial intelligence is so smart that it's going to take over and kill us. It's that it's really, really dumb and it already has taken over. So you get this range of views.

I don't doubt what you're saying, which is our discomfort and even our invocation of these very serious, extreme scenarios, might have to do with our discomfort that is encroaching on being a person. But I guess what I'm asking you is, as a literal existential matter, are you

concerned that artificial intelligence will ever go to a point where it could pose an existential threat to people?

DR. SUKHATME: In the sense of obliterations due to weapons, yes. But in the sense of the pervasive creep and taking you over and making you a parasite of the machine, blah blah, no. And in the extreme sense of, Are we in the Matrix? absolutely not. I don't worry at all about things like that.

MR. CALO: All right. Stuart, yes.

DR. RUSSELL: I don't know how much time we have. I could go on for the next 45 minutes.

So I think there are two ways that things could go south. One is a very gentle way, the gradually increasing dependency. I think a very good prediction of that comes from E.M. Foster, who normally wrote costume dramas, but he wrote a science fiction story in 1909 called *The Machine Stops*, where they have the internet, they have iPads, they have video chat, they have MOOCs. The machine does everything for you in terms of providing food and entertainment and communication, and so on. The entire human race becomes completely dependent on this existence.

And then there's the de-skilling. So he's really incredibly prescient. People lose real technical understanding of how the whole system really works, and they are no longer able to keep it running. So the human race ends up at a dead end. So I really recommend that. It's freely downloadable on the web, because it's out of copyright. So that's one possibility.

The other possibility -- I think Hollywood gets it wrong. Invariably, in the Hollywood scenario, it's somehow that the machine just spontaneously becomes conscious and therefore malevolent. I'm not quite sure why it's therefore, but it always is. I think that really misses the point. The reason why people like Elon Musk are concerned, and his metaphor is summoning the demons, meaning that calling on some very powerful technology and assuming that you will be able to retain control over it.

And it's summoning the demons, or the genie, or King Midas, or whatever. The problem is that we're not very good at specifying the objective. So King Midas said, I want everything I touch to turn to gold. The genie, or the robot, or whoever it was who was granting his wishes, should have said, No, you don't really mean that, do you? What you really mean is that, you say 'Abracadabra' and point to something, and I'll turn it to gold. That's a much better way, otherwise, you're going to die because your food and drink will turn to gold and that won't be good.

So what people are concerned about is that we

create incredibly capable AI systems optimizers where given some objective, the solution that they find -- by definition, if they're going to be more intelligent than us, it's going to be something we didn't think of. So not only did we not think of it, but we didn't anticipate that we would be very unhappy with that consequence. And at the moment, we have absolutely no science of how to specify objectives such that we are guaranteed to be happy with the optimal solution for that objective. All of the sciences that study rationality, optimization, and so on, assume that the objective is just specified exogenously. And it's the goal of the science, or in this case, the goal of the machine, to figure out how to achieve that objective in an optimal way.

To me, that's a fundamental mistake, and it's a mistake that I'm partially responsible for because that was the definition around the textbook on AI is built. What we need instead, actually, is a way of thinking about AI that's proven to be beneficial, meaning that whatever our objective truly is, whatever actually makes us unhappy or happy, the machine is guaranteed to behave in such a way that we are happy with what it does, that we are better off with the machine than without.

And that's a different problem, right? And in

particular, it means that machines should never take the objectives that we give them literally. They should always say, No, you don't really mean that.

You can boil it down to three simple principles. Number one, the machine's objectives should be to optimize our payoff. Number two, the machine doesn't know what it is. And number three, the machine has the ability to find out, primarily from observing our behavior, that what we do -- in fact, the entire record of what we've ever done provides a huge amount of evidence of what people want and don't want; what makes us happy and unhappy.

So this is just the beginning, but we can actually build machines that make us provably happy, and that we are probably better off with a machine than without. But there's a huge amount of science to do, not least psychology, because when you think about human behavior, we ourselves are far from rational. It's not even clear that we actually have objectives or payoffs. In fact, our behavior is the result of multiple competing subsystems within our brains, and none of them are close to being rational. So somehow you've got to invert through all that to figure out what we really want, and then have the machines act on our behalf to help us get it.

One of the things we might really want is

autonomy, which might mean that the right things for machines to do in the long run is to fade into the background; to provide scaffolding to facilitate our freedom our action, but actually not to do all that much in terms of meeting all of our needs; leave it to us to do a lot of that work.

MR. CALO: We are going to open it up to questions and start with this gentleman here. But before, I just wanted to summarize what Stuart is saying is that the AI may kill us. If it does, it's partially Stuart's fault, but he's working on it. So that's my summary of Stuart.

DR. FINKELSTEIN: I am Bob Finkelstein, Robotic Technology.

I'd just like to make clear that the commercially available driverless vehicles, driverless cars, trucks, etc., by 2020 or so will potentially be a weapon of mass destruction. They can carry payloads, chemical, bio, radiological -- or just the vehicle itself. A large truck in Nice, France mowed down dozens of people, killed them. They had an incident here, in this country, a car; in both cases, the vehicles were stopped because drivers were shot to death. And there won't be any driver, or there will be multiple of these vehicles operating simultaneously.

In any event, we have lone wolves, we have

organizations, organizational terrorism. They will be readily available. So my question is to Douglas: is his organization looking at the potential countermeasures? We've started to look at potential countermeasures, some of which would require subsystems by the manufacturer to be installed in the vehicle. So that effort towards countermeasures has to really be started, or examined, now and not after the vehicles are being sold in showrooms.

So anyway, I'm interested in if anything is being done at this point.

DR. MAUGHAN: I can tell you that at least in two areas of interest that we've talked at least about today, I already mentioned the fact that there is a fair amount of work going on looking at the counter-UAS (unmanned aircraft systems), counter-UAV (unmanned aerial vehicle) types of technologies for drones and how they might be used.

In the automotive space, we have a working relationship with a number of auto manufacturers, original equipment manufacturers (OEMs). That is a concern that they have as they look at building cars that are going to be autonomous, because in the end the car is just software. How am I going to update the software in that car? And if I'm a bad guy, and I can figure out how to update the software in the car, I now turn it into a weapon, right?

So I can tell you the auto manufacturers, that's their number one problem is how do I securely update the software in my automobile, whether it's with a driver or without a driver. It's their number one concern at the moment.

MR. CALO: Any other less terrifying questions from the audience? Laurel? Is it terrifying? Don't --

DR. RIEK: It's probably terrifying.

MR. CALO: Okay, all right. Fair enough.

DR. RIEK: I have a kind of follow up question for the individual from DHS. I'm just wondering -- so there's of course software security, right, that we've heard which is really important. But what about physical world hacking, to thwart sensors?

So we're roboticists, we know about the fact that sensors are really fallible and we can't have 100 percent perfect data. One could very easily project a picture of a brick wall, and make the car stop. It's really easy for one to physically hack the environment and thwart the sensors.

In the robotics community, we don't have any ways of dealing with that, so I'm wondering how DHS is thinking about this problem of physical world hacking.

> DR. MAUGHAN: You said in what community? DR. RIEK: I am wondering if in your community,

how people are thinking about physical world hacking. So take autonomous vehicles for example, or drones - [how will DHS prevent hackers from] altering the perception of these systems by altering the physical world?

DR. MAUGHAN: So in fact, I can't say that that's happening in the automotive space, but certainly I know in the drone space, that is one of the conversations that we've been having. Can I disable the drone by somehow -obviously, it's got something bad in it that's making it do something against us; can I modify that in a real-time thing to change the drone's behavior?

DR. RIEK: (comment off mic)

DR. MAUGHAN: Agreed. I was thinking from the standpoint of I can actually change the device itself; so that forces its behavior to be different.

As far as physical hacking, I am unaware that we have done anything in that space as far as making the environment look different.

MR. CALO: Well there's the GPS spoofing for drones, right? Making them land because they think that they're higher than -- but that's not what you mean though. You mean changing their world.

DR. RIEK: So what I mean is, so one of my colleagues does a lot of work on thwarting deep learning

systems. So what we can actually do is we can put up a picture of a brick wall and make an autonomous vehicle stop. We can project different colors of lights. There are a lot of really simple, low-level possible attack vectors. It doesn't take a PhD to hack the system by altering the physical environment to thwart the sensing systems on these devices.

This is a really critical security issue that would be good for DHS to think about.

DR. MAUGHAN: Agreed.

MR. CALO: Question in the front.

DR. POLSKI: Margaret Polski, Krasnow Institute.

So I want to follow on that a little bit, in terms of pushing toward research and what's the research agenda here? I think these are the things that I have heard you mentioned that we need to pursue -- catastrophic risk and our ability to estimate that, including the weather; data, big data, and the tools to capture, manage, analyze the whole works; deterrents and countermeasures, the item that was just mentioned; and then understanding machines better, how do machines think? What are their biases? What gets in their way?

And something that I didn't hear you specifically talk about, but I wonder if you would talk about and that is, our policies now, particularly in the DoD since so much of this discussion is focused on AI as a war-fighting tool, our policy is that humans will always remain in the loop. However, as machines become smarter, the loop starts to shorten, and command and control becomes more difficult. The period of time that one can exercise this control exists.

So if you could speak to that, and if there are any other research agenda items you might want to add to this list --

MR. CALO: What should we research? Gaurav, are we going to start with you?

DR. SUKHATME: So I think the third or fourth point that you mentioned, which is developing human understanding of what the systems are doing, I think, is crucial. It's a completely nontrivial thing. It's always easy to give a zero authority explanation for any technology, you simply call it magic, right?

That's trivial. And that's roughly where we begin. So when you're three years old, it's relatively easy. And it's difficult for us to appreciate, but for millennia, that's pretty much the state of explanation for everything in the world. That's what it was; that's what passed for explanations. It's only very recently that we

believed that you can give better explanations, because this is what science did, to change the way we gave explanations.

If you take an autonomous car, you can say that this is a system that drives itself, so you don't need to drive it. It's probably not a good explanation. You probably need something a little more sophisticated than that. Because what doesn't say, it doesn't say something about how that system would fail, and it will inevitably fail. It doesn't matter whether it's eight-ninths or nineninths.

In this sense, I think the bar is very high, because we know that people will also occasionally get into accidents. We don't actually feel the need to explain that in any way. The explanation is they're people. But we have to do something better with software that drives things like cars. And that is, I think, genuinely difficult to do. It's difficult because the explanation is designed for people. That doesn't make it any easier; it's not designed for machines. I think there is a vast space there on how you expressively and in some meaningful way get at the heart of what these machines do that humans will accept as a rational and reasonable explanation.

MR. CALO: A related point that I hear a lot in

these conversations has to do with how much safer than a person should a machine be before we deploy it. And it matters in part because machines may do things that people don't do in some instances. And we'd have to be really comfortable that the machine is a lot safer before we will tolerate these strange things that people wouldn't do. And I think that's a related point.

Stuart, did you want to respond, too?

DR. RUSSELL: Yes. No, I think there could be very serious backlashes after certain kinds of mistakes, which might mean running over a small child, or accidentally driving into a supermarket, even if, as Ryan says, they are on average much safer. I think that companies that go out, try to get ahead of the curve and deploy technologies that aren't really ready are really shooting themselves and everybody else in the foot. So I welcome this new partnership on AI, which is partly aimed at keeping everyone in line and making sure that no one spoils it for everybody else by making some serious mistake.

I think your point about the speed of decision making required in military context, the fact that the loop that the human is supposed to be in is gradually squeezing tighter and tighter. I think that's exactly correct. I

think there is a serious problem. It's a slippery slope, because what people are talking about now, Bob Work (deputy secretary of the Department of Defense) for example, is human/machine teaming, where you might have one human pilot and five autonomous fighter aircraft that are following him and protecting him.

But inevitably, if they get into a dogfight, well the autonomous aircraft are going to be attacking a humanpiloted aircraft from the other country and then those human piloted aircraft aren't going to survive. I think humans have no chance -- when you are fighting against a machine whose reaction time is a million times faster than yours, it can't last very long. So then you've got this very strong incentive to take the human out of the loop altogether and just go to full autonomy.

I think people have to think very carefully. Is there any stable point between where we are now, where all fighter aircraft are human-piloted, and full autonomy? And I'm not sure if there is a stable point in between.

DR. MAUGHAN: I would like to add to that, out of a military context and into just a defensive context in cyber security. So in today's world, everything is so much faster. If we always have a human in the loop, we never catch up to the bad guy online. So you really have to think

about where do I actually really need a human, because in fact what I really need is machine to machine, which is all autonomous AI types of capabilities. I really actually don't want a human until I'm asking them the question that says, Do you want to push that red button? Other than that, I don't want the human in the loop.

MR. CALO: Here is our plan for the next little bit here. I think we're going to take a couple more questions, because people seem really interested and invested in getting their question in the queue, which is great. And then we're going to move immediately into setting up the next panel, because that helps one of our speakers manage a time commitment that he needs to manage.

And so we'll take a couple more questions, and we'll move right into the next panel. Obviously, people, if they need to step out, they should, but there will be a formal break after the next panel.

All right? I think I see a gentleman right there with his hand up.

DR. BUTT: Thank you. Yousaf Butt, from the State Department. I have a question about the legal aspects of autonomous weapons. In regular warfare, you could find if a combatant injured civilians and he was found at fault, there could be a court martial process and you could hold them accountable.

In a scenario with autonomous weapons, who exactly is accountable? The person who wrote the code? The person who trained the algorithm? And not just in a military context, but also with vehicles, obviously; if there's an accident, surely the company there -- it's a little bit clearer that the company could be liable, but is the actual programmer or the trainer of the AI algorithm?

So in both the civilian and military contexts, if you could talk about the legal aspect.

MR. CALO: I am probably pretty well positioned to talk about that.

Basically, the cases that we've seen to date have in some ways really strived to find a person in the loop to blame. There are a number of cases where say a plane both was in autonomous mode, but also was improperly loaded in terms of where the load was placed inside the plane, where it crashed. And the regulators didn't even look at the autonomous mode, they just said this is obviously a problem because it was improperly loaded.

There was an old case involving an individual who put his plane in autonomous mode -- this was years and years ago, in the 50s, so basically it was a mechanical autonomous mode, autopilot. And it crashed into somebody.

And the court said, Well, this person should have been monitoring.

Increasingly it's going to be difficult to do. And the reason it's going to be difficult to do is because both criminal law and tort assume respectively that you mean to do the things that you do, or at least you should have foreseen those consequences. So when you have systems that behave in ways like Peter's robot that I mentioned this morning and was setting its own goals, or they just behave in ways that they think is the right way to behave, but no human foresaw it -- then you start to have difficulties. In the military context, this could come up either because a weapon system did something that no one intended, and therefore it's hard to say, Oh well, you meant to do this crime.

But it's also simpler still. I remember asking one of my colleagues who does maritime conflict, it is what he specializes in, and he gave me this amazing example of the idea -- and I know you work autonomous subs, Gaurav, maybe you could jump in on this too. He was talking about how in order to have the full panoply of rights in open waters, that some flag nation has to have effective control over that vessel. Which is generally accomplished with a person; we don't have to understand, we don't have to

explain it. We just say it's the captain, and the captain has got to report back to some flag nation.

But what we're doing right now is we're parking these autonomous military submarines on the continental shelf outside of China. And they're saying things like, according to my colleague, You're not maintaining effective control over this because these things are running autonomous missions. And the United States would say, Well no, we're the flagged nation. There's conflict over even whether there's effective control, which then has consequences.

So the idea that there would be this emergent behavior that human beings don't foresee -- my position is that it challenges a number of different areas of law, and not just liability.

I don't know if other folks want to react to that question. It's a bit of an area of specialty.

DR. RUSSELL: I think it is roughly right, that the person -- there is going to be a person who specifies the mission to the autonomous weapon. If they don't understand well enough what that weapon is capable of and the kinds of ways it can go wrong, then it will be like someone firing a heat-seeking missile randomly at some city and hoping that it hits just a chimney but of course it

might hit something else, like the back of a bus. That would be reckless.

I don't think this issue of accountability is going to be a wedge in which we can drive through a treaty. I think accountability questions can be resolved, and that's not the reason to try to have a treaty behind autonomous weapons.

MR. CALO: We had another question in the back, and then we'll set up for the next panel.

MR. WALES: This question is primarily for Doug, but I'd like the others to respond as well.

I appreciate the comment of applying formal methods to software development, and I think that could be a substantial mitigating factor in some of the concerns over autonomous vehicles. I think you'd probably agree that it's necessary but not sufficient, and I think you'd probably also agree that in information security, these formal methods haven't done much because we have a lot of insecure systems, and formal methods haven't really permeated far beyond academia and government.

So I guess I propose to ask what would be the policy initiatives and so forth that would make formal methods more interesting to the commercial world? Is it the liability and tort model? Is it something else? How do we

get people to build better software such that the general populace can have greater faith in the reliability or predictability in the wave of autonomous vehicles?

DR. MAUGHAN: Everything you said up front was exactly right. Formal methods has been on the back burner, the back back burner for decades. I think there have been a couple of really good examples, so the DARPA program, and the Cyber Grand Challenge that DARPA ran this last year was also a really interesting example of self-learning software that could protect itself while doing other things.

In the end, I believe it's still going to come down to whether you want to call it software liability, which today we all buy software and we take it as a -- the EULA (end user license agreement) says you take it as-is; you don't get any protection. In large measure, we do some of the same kinds of things on these critical systems.

And I think it's going to really come down to regulation or some type of thing that's going to force these autonomous systems to be better. I don't know how to define better in that case. Some of it might be formal methods. Some of it might be something else. But in the end, I think that's always been out sticking point -nobody is held responsible for the software, whether it's good or bad. And we've got to get to that position if we're
going to rely more and more on these types of things.

MR. CALO: Yes, I would -- go ahead, please.

DR. RUSSELL: Just very briefly. I think there are success stories. Airbus software is formally verified, so it can be done. But I think you need a forcing function, which is some combination of liability, regulation, insurance, and people insisting on guarantees. And I think self-driving cars are going to be one area where I am probably going to want guarantees; the smart home is another, and I think that this is something that people are maybe not seeing. It's not as visible as a self-driving car, but it's rapidly occurring. There are companies installing AI software into thousands and thousands of apartments right now, where it's got cameras. It controls your heating, your lighting, your electricity, your cooking. If I'm going to do that, I want guarantees that naked pictures of me in my bathroom are not going to be spread all over the web, etc. I think people will demand this, and they should be demanding it already.

MR. CALO: I think that is a great place to stop, for many reasons. But I just want to say thank you for a great panel. I really appreciate it. Please join me in thanking this panel for their thoughts.

Agenda Item: Panel 4: International Session: policies around the world

DR. ZYSMAN: My name is John Zysman. I will be moderating this panel. We are going to flow directly into this panel. We have been very focused to this point, really, on an American advantage on this set of issues. And I think what we want to do now is try and see the extent to which and the ways in which the advantages, the concerns, and the formulations of these questions are different from different national vantages.

We're really going to have folks with three different foci, Dr. Kenji Kushida from Stanford, who is going to focus on Asia and Japan in particular; Dr. Robin Mishra, who is a computer scientist, actually -- a lawyer, but with the German embassy here in handling a number of the technology issues; and in absentia, Prof. Dan Breznitz at the University of Toronto, where Gail has graciously made an exception to the rule that you have to be present and can't come in by satellite, as it were. Dan, who is Israeli; his wife has a family emergency in Israel, and so he is home with two kids. And bringing them to Washington for his talk seemed a little bit exaggerated.

That said, before we get started, I wanted to make a couple of remarks. This panel, this program, that

Gail runs and her colleagues run, really began about the time that the program that I run called the Berkeley Roundtable on the International Economy started out in California. It started with help from a number of people --Bob Noyes who, as some of you know, founded Intel; Jerry Sanders, Charlie Sporck, Steve Jobs, and help from a number of people who are here, including folks from IBM back then.

And all of those folks really believed they were creating a utopia. They dismissed in a real sense the concerns of folks like Kurt Vonnegut's book Player Piano. And what we hear from Martin Ford, what we've heard, really, in different ways, from others today, is that some version of Kurt Vonnegut's dystopia in the novel Player Piano is really upon us, because if you don't know the novel, it's a world in which the only people who have jobs are all of us, and everyone else is out either in the army or in the wreck crew cleaning up after the city and keeping things in some version of order. The novel came before semiconductors, so they were all giant mainframe computers, but the story was the same.

And now the question really is are we really going to have that utopia or that dystopia? The dystopian view we've heard loud and clear here, and I wanted to say a couple of remarks about that before turning to my

colleagues here.

The first is that the emphasis on the destruction of jobs clearly ignores two things. One is what jobs will be created and in what form, and two what extent will technology transform the character of work in its very real sense?

That transformation depends on partly how tasks are recombined, and what kind of human complementarities can be found for the work that is to be done. Many of the numbers that we've seen, the Frey and Osborne study that is so widely cited -- it's very interesting. It emphasizes 47 percent of tasks, of jobs, are automatable. The OECD numbers suggest, as was mentioned, nine percent.

Now I've not really taken a close look, and I haven't gotten inside of the OECD numbers, as to what is this radical difference between the number of tasks that are going to be displaced in the two different studies. The implications are massive, obviously, and so one thing we need to do is think about that.

But this issue of transformation very much reminds me of back when we at the Berkeley Roundtable started looking at robotics back years and years ago. And what was clear was that the American automobile companies were looking at substituting robots on a one to one basis

on the assembly line, and in fact really had no thoughts at all about how you could recombine or reorganize production. Machine tools, not simply robotics, were really employed and deployed by Japanese companies, particularly Toyota, in what became Lean production systems, and a radically different way of organizing production that gave them massive advantage.

So the question exactly of what the imagination is going to be about how these technologies are deployed, how they are recombined is absolutely critical. And one part of that of course becomes the issue of user interfaces. Someone was remarking earlier, I think it was Ryan, about clerks and the displacement of clerks in the old days. Well the biggest displacement of clerks and typists is Word, and it's a user interface.

I recall actually looking at Italian machine tool companies, 25 years ago, when they deployed a BASIC -literally, the old computer programs of BASIC -- on top of C++ so that their workers could in fact have control of the factory because they thought that the workers had a great deal to contribute. So part of the issue of how technologies are transformed or recombined really will become an issue of how we look at work and workers; to what extent are they simply costs, and to what extent are they

assets and what sort of assets do they actually represent. What kind of assets do human capabilities actually represent, and in this world how do they get recombined?

That said, I think that what's particularly interesting to me about both what I know of the German conversation, and what I know of the Japanese conversation is that they are in important ways different from the American discussion. Whether or not the outcomes are the same, whether or not -- I think we want to hear something about that from the Israeli conversation.

So my view is that even if Martin is right and my colleague Stuart Russell is right, and that we are, in some sense, doomed, we have to make an effort to try to build a world built around human capabilities, and that the question is we have to speculate and try and imagine what that would be like and try and make it happen. If we fail, we fail. But we have to make that effort.

Now, we were going to let Dan Breznitz go first, simply because he's calling in from Toronto, and we want to let him be able to go back to taking care of his kids.

DR. BREZNITZ: I'm here, if you hear me.

DR. ZYSMAN: Why don't we start with you, and you can begin whenever you're ready.

Artificial Intelligence and Robotics: Thoughts, Reflection on Policy Inaction in Israel and some nasty general predictions Dan Breznitz



Presentation of Dan Breznitz, 12/12/2016

DR. BREZNITZ: Sure. Learning almost like a machine, learning from past ambitious attempts to do this online, a very sophisticated presentation, what I decided to do instead is doing an extremely simple one, basically to open questions. Some of them I just heard, so some of my questions were just discussed at length, but some I don't think were. I hope this would then allow us to open conversation with you continuing into the future.

So what I'm going to do is very quickly, just let you know what I think is the situation is in Israel, as an ex-Israeli who hasn't lived in the country for 20-something years, so what I see when I come to visit and do some research; talk about some nasty general predictions, which I have not heard enough about, assuming that many of those things are well-known to all of you, and move on.

So can I have the next slide please?



Presentation of Dan Breznitz, 12/12/2016

So in Israel, you have -- the discussion points are going to reflect this -- Israel actually had a very long history of growth, of artificial intelligence and robotics, especially around security. And very recently, because Israel actually had an active policy setting of those, but very recently in the last few years, especially in the last 12 months, a complete deadlock.

I will then talk about what are the real trends that we see in Israel and outside Israel. Assume for a moment as a political economist, the people who do that in the market actually look to maximize their profits, and say that if I'm correct with that, and I assume I am, this is going to be politically explosive compared to which Uber and Airbnb are nothing. Can I have the next slide please?

Israel: a Long History of Al (dual-use) "Self-Driving Cars? I solved that 30 years ago – but a faster car, actively hostile conditions, and a load that went boom" "I was on sabbatical in Texas in the early 1990s, and I decided that I really, but really hate mowing the grass, and I thought to myself I am probably not alone – so I took what I knew from the air force mission planning about self learning navigating machines... by 1995 the company started selling worldwide" The same goes to even more sinister things: system that learn to recognize imminent attacks by changes in flows of data and troops; and systems on the battle field – questionable how much Al-autonomy you really want these "smart" machines to have



© copyright 2016 by the author

Presentation of Dan Breznitz, 12/12/2016

So a very large history of artificial

intelligence, dual use, almost classic. So this a quote from an interview I had where I asked people about selfdriving cars and Uber. Basically, the argument that they have been giving me is that they have been dealing with issues that are very, very similar; they just used to call them smart bombs. And again, that might be correct, but you can see where those things are coming from.

This also, and the fact of how long a history it is, is that this a company which is now 21, 22 years old that has basically developed robotics for the use at home and in commercial spaces, doing things that were supposed to be done by humans. So the first mowers, and then the first hoovers. It used to be called Friendly Robotics, now I think they're concentrating only on the mowers.

And again, what you see is researchers that were deeply into the security who then faced tasks that they hated, in this particular case, and opted to use them with what now we will see as extremely primitive computing. So their ability to do that now, especially assuming that you have connection to unlimited computing maybe via the cloud, is exponentially bigger. We should assume that those things will happen even to things which we think are almost irreplaceable at the low end.

The same, however, goes and you've just had a whole panel of that, but still goes to even more sinister things. And it's lovely to talk about aircrafts, but let's

talk for a moment about the security dilemma of Israel, which is an extremely small country. So if you have a surprise attack against Israel which takes an hour before Israel reacts, there is no country.

So there has been, for at least 25 years, attempts to use artificial intelligence, what we now call artificial intelligence, to create systems that can learn and recognize imminent attacks by small changes to the flow of data, to the movement of troops, to just where you locate various equipment. And of course, there are all those autonomous systems on the battlefield which we just talked about.

I heard about the red button. The question I think that we need to ask is are we close to a point where those systems, if you want them to be efficient, actually cannot even wait for the red button? Meaning, for example in the case of the Middle East or somewhere else, assuming you trust your system, and assuming again that it's must faster than a human being, can it decide that someone has declared a war on you in a surprise attack and start mobilizing before you make a phone call, even to the president?

Any one of you who read the 9/11 report would know that this is not just a hypothetical question. And

those are the questions -- not just the red button, but to decide that you're under some kind of an attack and therefore you need to mobilize your system. How much autonomy do you want those systems to have? The rest, I think, from what I heard from the past panel, it's almost all of those questions that you just have dealt with at length. Next slide, please.



Presentation of Dan Breznitz, 12/12/2016

So under those conditions, and especially with so much of that coming from the military, or the security, you would expect Israel to be actively developing policy frameworks of how to deal with that, especially since as some of you know, Israel has moved to extreme high level of inequality. So questions that have to do with the labor market are prominent; indeed Uber is not legal in Israel. The Minister of Transportation has prevented Uber from coming, which actually allowed an Israeli software company to create the app called Gett. It used to be called GetTaxi, but now it's called Gett with a double T, which is one of the biggest competitions to Uber, but using real taxis.

So you would assume that in this kind of an environment, and especially since high tech is the only sector that actually is working in Israel, you would have discussion. The problem is that the main organization, the Office of the Chief Scientist, the main innovation development agency is for those matters basically has been at a halting point for the last year or so, because it is going through its most significant reorganization since its inception, for very good reasons, and that is its mission was to maximize civilian R&D. If you look at the last ten years, Israel has maximized civilian R&D, the highest in the world by far. Therefore, the real question is, what should be the role of government in innovation in the country? A decision was made to change the legal structure of OCS.

However, while you do that, you have inability to come up with new policies. To make it more complex at the same time, the Office of the Prime Minister opened a center for Cyber, which is more oriented to cybersecurity, but also is sort of muffling its way to a certain degree into this and to create even more fog of war about what is allowed/not allowed, what is exportable/not exportable from Israel.

And speaking about the Prime Minister, he -- the person -- is now also, for political reasons, the Minister of a large number of ministries, including the economy, which again means that even in the political level, you have a complete deadlock. Because while he might be a really capable person, no one can really control 15 different ministries, including being the Prime Minister, and make all those decisions at the same time.

So in short, you have a deadlock and inaction that, to be honest, put DC's to shame in its high level of ineffectiveness. And the result is what really decides what happens in Israel is decisions by entrepreneurs and investors -- meaning the market, and where people think they can make money. Next slide please.



Presentation of Dan Breznitz, 12/12/2016

Which leads us to general issues, which I think are more important than what happens in Israel because Israel to a large extent, basically imitates Silicon Valley and New York as to where they think they can have money. So first of all, artificial intelligence and robotics are to supposedly quickly replace humans in what we call low or mid-skilled level jobs, including things we didn't think they would do, maybe like self-driving cars.

And driving and manufacturing seem to have captured the public eyes, and again, even in Israel, Uber is illegal. But if you really analyze what we can now do with artificial intelligence, the real bucks, the real money to be made is not in replacing low-paid workers. And if you look at manufacturing, if you look at the cost of labor in overall manufacturing, it's miniscule. It's really in replacing high-value humans with manipulated data; a lot of them in services, but not all. So research background analysis, to finance sports scouting -- if anyone of you like sports, just look at what people now do, scouting, eager to figure out who is a good player to buy and also to decide what should be your game plan against a specific team. Probably many doctors, again, who needs a banker if AI can do it for better and cheaper, and so on and so forth. Next slide, please.

General Issues: Why is this politically explosive

- Al will soon disrupt not the weak unorganized and unloved swaths of humanity that most politicians find easy to disregard (at their own peril) BUT the modern nobility of our capitalistic Western Societies:
 - Lawyers
 - Doctors
 - Bankers
 - Engineers

Many of these are already well organized and politically connected, do not expect them to submit to become "mere" humans without a fight



INNOVATION POLICY LAB at the MUNK SCHOOL OF GLOBAL AFFAIRS SCHOOL GLOBAL AFFARS

© copyright 2016 by the author

Presentation of Dan Breznitz, 12/12/2016

So if you put for a moment your political hat on, that means with artificial intelligence, it will disrupt not only the weak, unorganized, and unloved slobs like taxi drivers -- who loves a taxi driver? -- of humanity that most politicians find easy to disregard, by the way at their own peril, as we've just seen in this election, but truly what I call the modern nobility of our capitalistic Western systems. And if you think about the professions that lead to positions of power and high status, it's lawyers, doctors, bankers, and engineers. If indeed what I hear and what I see is correct, for example, for many cases where you want to have background on the landscape of the law, you can get a better one by software instead of by high-paid lawyers; this is going to have a significant disruption both on those professions and also on the ability by the way, which is another issue of us to create the true master of our profession, which we might want even if we have AI. Because if you don't have those low-level in which they those learn those skills and really immerse themselves in the data, we will not have really successful IP lawyers for example.

But the real issue is that as soon as this

disruption would be evident, if you look at those professions and those people, not only are they very well organized, they are small in numbers so they can organize quickly, but they also have most of the political power and money and the channels of influence. And they will not let the market rule and see which one of them and their sons and daughters are going to be unemployed. They are probably going to be extremely active in reshaping the arena. And if we do not develop at least a set of principles of how this should be dealt with, we should assume that the people with the most political power will actually make the decisions about the trajectory of technology.

And with that, I'll stop, hoping that you actually heard me because I hear nothing back.

DR. ZYSMAN: Thank you, Danny. I hope you are hearing our thanks, and we will come back if you can stay online, if the two kids will leave you be for a little bit. With that, let me turn to Robin, let's just go down this, who will discuss the German case.

DR. MISHRA: Okay, does it work? Yes. So thank you very much for having me here.



Robotics and Artificial Intelligence: Policy Implications for the Next Decade

- International Panel -

Dr. Robin Mishra



Presentation of Robin Mishra, 12/12/2016

As I told John, I am not a data scientist, but a lawyer, but let me give you some perspectives about the way Germany is going. I have some restrictions, given by John, who told me to be optimistic, which is not easy after Stuart's remarks and what we heard recently, but I will really try.

My basic line is Germans have a thing for science fiction, and they think about robots as they do in science fiction novels. But what's happening on the ground is somewhat different, and it doesn't scare people so much. So I will try to cover these issues.



Dr. Robin Mishra

The National Academies of Sciences

December 12, 2016

Presentation of Robin Mishra, 12/12/2016

So there's one term in Germany, which is called Industry 4.0, and I'll talk about that later, that is kind of the basic term for the changes that we experience at the moment in our engineering sector and in our industries. And as it is called Industry 4.0, I will talk about TV Thrillers 4.0, Research 4.0 and also about Work 4.0, which is the issue that Dan brought up. Next slide.



"Tatort" - German Crime Series

→ TV Market Share (2015): 25,8 %

Tatort Stuttgart "HAL" on August 28, 2016



Dr. Robin Mishra

The National Academies of Sciences

December 12, 2016

Presentation of Robin Mishra, 12/12/2016

So there is a TV crime series in Germany, and I will start with that because something happened there. We had two thrillers in the last month that were dealing with the same issue, an AI creature, you see the AI creature called Hal on the left, a picture. And they were, as you can imagine, the murderers in two cases. So in this case the guy who designed the AI, who you see on the left side, he wanted to stop his artificial intelligence, because he realized that it went too quickly and didn't work the way he wanted it to, and the AI wanted to replace him on the job and so on. So he tried the delete the AI, but it wasn't possible, and so the AI decided to put him in a situation where he was killed. This was the first example. Next slide.



Tatort Bremen "Echolot" on October 30, 2016



Dr. Robin Mishra

The National Academies of Sciences

December 12, 2010

Presentation of Robin Mishra, 12/12/2016

This happened a few months later. A woman working in the AI industry had a car accident, she was killed. But as you can see on the right picture, she was still alive. So her daughter talked to her via her iPhone. The police investigated the case with her. And at the end it turned out that she wanted to leave the company, and so they decided, the artificial intelligence decided to manipulate the car.

So that's what a lot of German people can see on TV. It's on prime time. About 25 percent of the German TV



watching audience has watched that. Next slide.

Presentation of Robin Mishra, 12/12/2016

If you watch what's happening on the ground, you see a much more diverse and intricate picture. Perhaps, you can skip the next two slides and come to the next one. We have different research institutions in Germany. And there is this German system of having organizations like Fraunhofer, who have an applied research facility, whereas Max Planck covers more basic science.



Fraunhofer Institute for Factory Operation and Automation IFF (Magdeburg)

- Research areas:
 - Digital Engineering & Industry 4.0
 - Convergent Supply Infrastructures
 - Smart Work Systems
 - Resource efficient production
 - make factories more energy efficient
 - reducing transportation
 - implementing smart energy cascades
 - closed energy & material cycles







Max Planck Institute for Intelligent Systems (Stuttgart, Tübingen)



Presentation of Robin Mishra, 12/12/2016

And there is a center in Germany which is really at the heart of German AI development, which is the German



Research Center for Artificial Intelligence.

Presentation of Robin Mishra, 12/12/2016

It's designed slightly differently from all the other research organizations. It's a public/private partnership, so it's funded by regions, as well as the federal government. But it also has industrial partners. German partners like BMW, Volkswagen, Deutsche Telekom, Bosch, but also American partners, for example, Google came in a few months ago, Intel, John Deere. And the industrial shareholders even hold a majority on the advisory board. So they have a certain right to talk about what the interesting points in research may be. Next slide please.



Presentation of Robin Mishra, 12/12/2016

So this is what we are talking about in Germany, at the core of our discussion. Industry 4.0, and as you can see it on the steps we talk about first, the Industrial Revolution and Mechanization; second, electrification; third, automation; and fourth, networking. That's already a different perspective from what we have heard the whole day. So it's not about something disruptive, something we have never seen before, but it's the fourth step in development that we have gone through, and we may be able to adapt to it.

I don't know if this assessment is right, but it gives you an impression that we are talking more about a

kind of incremental change or another step. And we have mastered the first three steps, so it should be possible to go on and move to the next one. Next slide, please.



Presentation of Robin Mishra, 12/12/2016

There is some economic potential. So you know the figures are always different. In Germany, there is a study saying we may have 430,000 new jobs. At the same site, 490,000 low skilled jobs may vanish from the industry. But there are some positive examples. I think the most wellknown one in Germany was that Adidas, the famous sports manufacturer, moved one of its production facilities from Indonesia back to Germany. And they will have a very specialized manufacturing -- you can order a shoe in a shop, and they will try to produce and deliver it the next day. You can really design it and talk about what it looks like. And this has brought back jobs to Germany. So everybody can see there may be an upside and not only downsides to development. Next slide please.

Enbassy ef the Tederal Republic of Germany Washington			
Global economic potential of the Inte	ernet o	f Things	
Nine settings where added value is expected	Size in 202 Low estimate	25, S trillion ¹ High estimate	
Factories - eg., operations management, predictive maintenance		1.2 - 3.7	
Cities - eg., public safety and health, traffic control		0.9 - 1.7	
Human - eg., monitoring and managing illness, improving wellness		0.2 - 1.6	
Retail - eg., self-checkout, smart customer-relationship		0.4 - 1.2	
Logistics - eg., logistics routing, autonomous vehicles, navigation		0.6 - 0.9	
Work sites - eg., operations management, equipment maintenance		0.2 - 0.9	
Vehicles - eg., condition-based maintenance, reduced insurance		0.2 - 0.7	
Homes – eg., energy management, safety and security		0.2 - 0.3	
Offices – eg., augmented reality for training	1	0.1 - 0.2	
*Adjusted to 2015 dollars, for sized applications only; includes consumer surplus. Numbers do not sum to total, because of rounding	I \$ 4 trillion -	\$ 11 trillion	
	Source: M	icKinsey Global Insti	tute analysis, June 2015
Robin Mishra The National Academies of Sciences		Decembe	r 12, 2010

Presentation of Robin Mishra, 12/12/2016

Some industries and factories are -- I think the most important ones, where you can expect quite a lot of added value. Next slide, please.



Presentation of Robin Mishra, 12/12/2016

So what are the consequences for the German economy? There are challenges. Germany has a strong exportoriented economy, which means that we really rely on certain industries. We rely on exports. We rely on trade, which is as you know, not an easy issue at the moment. We also rely on industries that are challenged, especially challenged by machine learning and AI technology: automotive industry, manufacturing, engineering.

And there is another thing, I think you heard about it in the law panel a little bit. There are data privacy laws, not only in Germany, but in also in Europe, and they may slow down machine learning. So there is for example on the European level, what is known as the right to be forgotten, which is from a data security point, not a bad thing. So if someone uses your data, you can ask him to give you back that data and not use it for anything else. That is not easy, as you can imagine, in an artificial intelligence system where a lot of data is processed, and it's not easy to get one set of data out of it.

On the other hand, as I said, there is a general feeling in Germany that the industry may be able to cope with all that. Germany has a strong industrial base, as you may know. So German industry still accounts for 30 percent of GDP, compared to 20 percent in the US and Great Britain and France. There seems to be a strength in incremental changes, which means that you really modify AI and machine learning tools to those specific industries, with very different outcomes. There is the structure of small/medium enterprises, family owned enterprises, the so-called hidden champions who know very well how their market works. While it's important that this market still exists, that's the big question; if it exists, they may be the first to occupy the new space.

People feel that we have a high quality research and university field. And there is this kind of famous vocational education and training system. It may be

possible to do the reskilling and the skilling on the

foundation of that. Next slide please.



Presentation of Robin Mishra, 12/12/2016

If you look at the individual company, it may not be that easy. So there is a feel that platforms and consolidating companies may take over some of the existing B2B models. There may be risk aversion in smaller/medium sized companies, which may lead to slow adoption of machine learning. There is a huge issue -- and John, we talked about this -- with data sharing. So small and medium enterprise know that the data set they have is extremely crucial for their business model.

But the real benefits of Industry 4.0 are only generated if you bring together data from different companies, and how you can do that, and how you can do that so that it's safe and secure and all that is a huge question. So the smaller and medium sized companies, they are really trying to keep their data because so far they don't know what's going to happen with it.

And of course, there is a limited size and scalability. So Germany, in terms of people, cannot compare with the US or China. And smaller companies may not be able to compete with the big companies in the engineering and automotive sectors.

Still, we feel that there are also opportunities for the individual companies. So they may minimize time for development, reduce downtimes. Predictive maintenance is a big issue. Smart sensors, John, you were talking about user interfaces; there may be interesting applications for that. And as some of the companies have the real production data, they feel that they may get a competitive advantage over companies that have data sets but may not be able to compare them with real production facilities. Next slide, please.



Presentation of Robin Mishra, 12/12/2016

The last issue is about work and that dovetails with Dan. We had a discussion in Germany which led to the distribution and publication of a so-called white paper and these findings are really new. So I'm happy that I can talk about it today. We were discussing what effects AI machine learning may have on jobs. Next slide please.



Challenges:

- Digital transformation → lead to unemployment (esp. for low skilled workers)
- Quick devaluation of qualifications
- Demographic Change

Opportunities:

- New jobs in services, healthcare, education etc.

Dr. Robin Mishra

The National Academies of Sciences

December 12, 2016

Presentation of Robin Mishra, 12/12/2016

So the challenges, of course, are that these

mechanisms, these tools may lead to unemployment,

especially for low-skilled workers. Next slide, please.



Presentation of Robin Mishra, 12/12/2016

So you can see that it's comparable between Germany and the US. You see, and I don't know if I really share Dan's perspective in this regard. You see that a PhD has much less probability of losing his or her job to automation than someone with only a primary education. So all these forecasts will have to be proven. But you see a strong connection between the skills of a person and his or her chances of being employed in the future. Next slide, please.



Challenges:

- Digital transformation → lead to unemployment (esp. for low skilled workers)
- Quick devaluation of qualifications
- Demographic Change

Opportunities:

New jobs in services, healthcare, education etc.

Dr. Robin Mishra

The National Academies of Sciences

December 12, 2016

Presentation of Robin Mishra, 12/12/2016

So first challenge, unemployment. Second challenge, quick devaluation of qualifications; I think we heard about that today. The demographic change, I think my colleague with a Japanese perspective will talk about that. In Germany we also have a shrinking population. It may not be easy to find the right people for the right jobs. But there are opportunities as well. There may be new jobs and services in health care and education. Next slide please.


Presentation of Robin Mishra, 12/12/2016

So if you look at the bottom, there are other business services, IT services, social services, health care. There are studies that show there might be more jobs in 2030 than there are now. We know that we are talking about a black box, but there are serious people who think there may be more jobs in the future. Next slide please.



Challenges:

- Digital transformation → lead to unemployment (esp. for low skilled workers)
- Quick devaluation of qualifications
- Demographic Change

Opportunities:

- New jobs in services, healthcare, education etc.
- New quality of learning on the job
- Demographic dividend

Dr. Robin Mishra

The National Academies of Sciences

December 12, 2016

Presentation of Robin Mishra, 12/12/2016

So new jobs. New quality of learning on the job. This is really very optimistic, but it may be possible that people who know that they have to change their qualifications, update their qualifications, may have more opportunities to do so. We talked about MOOCs before today; there are other ways of upskilling and reskilling that may be easier in the future. It may even be more interesting for someone in the workforce who has new incentives to work on new issues.

While there is demographic change, there is also conceivably a demographic dividend. I don't know who said

it, for example, that you shouldn't do things that a machine can do. So there may be really hard jobs that you won't miss in the workforce, and people may have the chance to do more sophisticated work on certain issues. Next slide please.



Tools:

- Strengthen digital literacy
- Strategy for lifelong qualification & upscilling (turning unemployment insurance into a labor insurance?)
- Incentives & benefits for founders and start-ups

Dr. Robin Mishra

The National Academies of Sciences

December 12, 2010

Presentation of Robin Mishra, 12/12/2016

In this white paper, there were some ideas of how we could do this transformation. This is only to give you an idea about what is being discussed in Germany. Strengthen digital literacy. May we have better strategies for life long qualifications. Upskilling, there's one idea to turn unemployment insurance into a kind of labor insurance, which would be available for the upskilling process, and would help you through all your working life. So this is very ambitious, but it's an idea. There may be more incentives and benefits for entrepreneurs and startups, which is something where the US is really better and more advanced compared with other countries, especially in Europe. Okay, next slide.



Presentation of Robin Mishra, 12/12/2016

So working conditions are changing. There are some challenges. You know about that. Excessive demands on employees. The boundaries between work and leisure are kind of fluid. Next slide please. I found it interesting that you really have figures where you see that evening work -- night work, Saturday work, Sunday work -- has become more and more important and that being responsible for two children, like Dan, is really an issue not only if you are talking to an audience like that. Okay, next slide please.



- getting rid of physically and psychologically demanding work
- sovereignty of working time and working place
- self-determined life planning

Tools:

- collective or company agreements with room for experimental spaces
- working time accounts

Dr. Robin Mishra	The National Academies of Sciences	December 12, 2010

Presentation of Robin Mishra, 12/12/2016

So there may be benefits like getting rid of physically and psychologically demanding work. There may be more control over working hours and one's workplace, and you may be more self-motivated if you are planning and organizing your life. There are some ideas, for example, work time accounts, I don't know if you've heard of it here. The way is works is, you collect time in earlier stages of your life and get it back later, when you can invest it in upskilling or a sabbatical or a change between business and academia, perhaps. There are different models that add value and are being discussed. Okay, next slide.



Follow @GermanyinUSA @MishraRob on Twitter

Dr. Robin Mishra

IIshra The National Academies of Sciences Decen

December 12, 2016

Presentation of Robin Mishra, 12/12/2016

Which is the last one. So what I wanted to point out is that there is a discussion in Germany, and there is a certain scenario of robots taking over. But on the other hand, there are also some very realistic thoughts on what machines are able to do now and how we may be able to integrate them into our structures, which have proven to be good in other situations like the social system, the strong SME community, in order to be able to cope with the change. If you watch that comic, many people in Germany will perhaps hope -- or pray, I don't know, after what I've heard today -- that this guy, this employer will not find someone, will not find a mechanical version of that person, but will really find at the end of the day that the person is the best guy he can get and will be even better for the employer than the machine which is driven by AI or machine learning.

So I will leave it at that, and I am looking forward to the discussion.

DR. ZYSMAN: Thank you very much. Before turning it over to Kenji, as an American observer for a long time of European political economy, I wanted to just emphasize several of the points that you've been making, Robin.

The first is that the German discussion is very much a corporatist discussion. That is, industry, university, unions, and labor. That was evident in the kind of slides that you were presenting. But I don't think it's automatically evident to an American audience, and I think it needs to emphasized. The original Industry 4.0 document and the Work 4.0 documents are both corporatists in that

very character.

And I was a couple of weeks ago in Munich at a German National Academy of Sciences and Engineering meeting on the future of manufacturing. And one of the things that struck me is that the Fraunhofer Institute that you were mentioning would emphasize that for Germany, the whole discussion of cyber is manufacturing to which cyber is added. They contrast it with the American vision of its being software that is now swallowing up all of manufacturing. It's two very different ways of understanding much of this.

The German story, for us, I think, is particularly important in the issues that you raised, in that the German economy is really built on skills, on a very highly skilled workforce. The question is, if that's displaced, it has nothing to sell. So this problem actually has to be solved in a very fundamental kind of way for Germany to preserve what it actually is.

And the issues that I heard, actually there, were as you emphasized this issue of data, the enormous concern of how to create interoperability and interconnection of small and middle sized companies and how in fact to allow their data to both be protected, since that's the basis of what many of these companies are, and at the same time shared, which in some ways is a contradiction. But in fact, for the Germans, it's clearly a very basic issue.

And the last point that I would emphasize, because I know that Kenji is going to raise it -- Dr. Kushida is going to raise it, and it's a way of handing over, is that as you mention of the shoe company where you come in and you can basically have your custom shoe designed and then manufactured, there are a number of those in a real sense, unlike a lot of the ways in which it's often seen in the US, the automation processes are in a way often seen as a mechanism of reinforcing this ability to main custom design and specialty product.

I know that Kenji, and it's a way of handing over to you, found that Japanese firms that heard Wolfgang Wahlster, who is one of the authors of Industry 4.0 was our guest in Berkeley. In fact, when he did a presentation, the Japanese companies were shocked by what they heard because it stood their model of production on its head.

And with that, Kenji.

DR. KUSHIDA: I do not have slides but it is a sleepy time of the day now, so I am going to stand up here and wave my hands a lot so that you can stay awake, even though it's a little dark here.

So first is a general point, which is a point

that I learned -- this is probably I have worked with John for so long and been indoctrinated, you can put it that way. But I believe it's true, that BRIE, the Berkeley Roundtable on the International Economy, a lot of vantages, it's not about the technology and tools themselves, but how they're implemented. And that's how it has impact on the economies and societies.

So how does technologies get implemented? Well, it's about the rules and regulations. And these matter a great deal.

So what sets the rules and regulations? Well, it's politics and political dynamics that shape how the rules and regulations, the trajectory of those, which then shake what is implemented and how it unfolds.

And this topic, we're right in the middle of that as it's unfolding, and I used to have -- I mean, this is the wrong crowd obviously, and the wrong place to say that, Look, politics matters because it can't be more obvious. But it's a good starting point.

And here's where Japan comes in, and it's been fascinating. Why do I know about Japan? I've had the wonderful fortune and the misfortune of being there eight times this times this past year, eight times, last year. One of the projects I'm doing at Stanford is bringing large

Japanese companies who would like to harness Silicon Valley innovation. So how does Silicon Valley work? What are their strengths? What are their weaknesses? Organizational challenges, how do they work? These kinds of things.

So I talked to a lot of big businesses, or a lot of the more interesting ones -- because selection bias, not interesting, I don't want to talk -- and then different parts of the government that are very interested in these things. What's most striking, what's very, very striking, is that the debates there are so very, very different at a very basic level from here. I'm not saying they're right. But I'm highlighting their differences.

Because I don't have slides, I'm going to talk in bullet points. So if you're taking notes either on your head or on a piece of paper, you can follow there.

So, punchline -- robotics and AI are seen as a solution to major labor shortages caused by the looming demographic crisis. People have been wringing their hands in Japan for two decades now, saying, Look, we have this postwar population bulge that is working its way through. They're going to retire, and then we have a population inverted pyramid. What are we going to do?

Now there's a solution/panacea. It's probably not exactly what it is, but they've latched on to robotics and

AI as one of the panaceas because it's accompanies by things like real rising wages in regional restaurants and other places because people won't work there. Some restaurant chains have to close their regional branches because they can't find people to work there.

Deflation is over in that area, and in other parts too, in areas such as elder care, high end manufacturing using skilled labor, agriculture, transportation. And to replenish declining pools of skilled labor -- these are the things I'm going to talk about a little more in detail. I'm not saying they're right. I'm saying this is what the political debates are saying that AI and robotics are fantastic for pouring money into them. We'll talk about where the money comes from and what the dynamics are in a second.

Agriculture. So 65 percent of Japanese farmers are over 65 years old. We saw from the Google presentation in the morning, the cucumbers example, and notice those people did not look like young farmers. That's typical; in fact, they're on the younger side. It's a big issue.

So what you cannot do in Japan, just geographically, is large-scale agriculture. We know that. So you do smaller scale, but more high-end. So if you've travelled throughout Asia -- I've been to Korea a couple of

times, I go to Hong Kong a bunch, I've gone to Singapore -high end fruits are often Japanese. They're airlifted in. The pears that damage easily, the nice peaches, etc., but these have to be done by people. The people are getting older.

Well, what if robotics and AI can be a solution to this? That's great because the amount of untended agricultural land in Japan doubled over the last decade. There's land, lots of land, it's really going to the wild, but from different dynamics than say, Detroit. And you watch the news, it's very different. And the people don't have guns. The issue is somebody got hurt. It makes news because the elderly person got run over by a boar that came out of untended land. It's seen as an actual issue; you see it within the first ten minutes of a 20 minute national news.

So agriculture, high-end products. Not scale, but quality. Well, let's apply deep learning to this, and then figure out mechanical solutions, because it can't grab the tree and shake it. You have to actually pick these things, and each Asian pear has a little wrapping on it that a person puts on. Now for all of this, maybe robotics has a solution. So let's pour money into agritech.

Second, elder care. With over 35 percent of the

population going to be over 65 within the next 70 years, if it's a linear projection -- linear projections are usually wrong, but Japan, that has a population of 127 million now, is projected by 2100 to have about 80 million. No advanced industrial economy has ever lost that many people. So these are usually, Let's get immigrants. No. The politics of Japan means that there will probably not be flows of immigrants. So the population will decline.

So elder care, a large proportion of the population will need elder care. There are several things. Information and coordination of needs among different caregiving and healthcare provision entities. Right, these things need to be coordinated. There's a need for that.

Information monitoring, and then of course, there's prediction before disasters. And then of course, there's quick response after a disaster. Most high-end jobs are in metropolitan areas, and most people that need the most care are not. So how do you do this?

Interestingly enough, and this highlights how Japan is a little different as well, a new service by an entity that can go check on people every day to make sure they're okay in rural areas. Who would you trust to do this? Of course you would trust the post office. So a new service by the Japan Post, which recently got privatized,

is to go check on people. They don't just deliver the mail; they go in and check to see that the elderly person that's usually there for the kids to buy the service to make sure they're okay, to check in every day.

That being said, that's a human, labor-intensive solution to this; weather can be bad, etc. -- so automate this. That's great. People are excited about this kind of thing. And the information privacy issues are of course a bottleneck, but the information need is going to be huge. The need for coordination is very large.

And remember, emotional and mental health. This is a country where the Sony Aibo, from 1999, that little dog -- unfortunately when Sony lost money from betting on 3D televisions, they stopped it. But people got very attached to their robotic dogs. And there have been loads of documentaries showing that after Sony discontinued the maintenance of these, that third party fixers showed up because people became very emotionally dependent on their robotic dogs. They'd take them for walks. Depending on how they interact with it over the years, if you're very excited with it, it will be a very excitable little dog. And if you're very chill and calm, it will be very chill and calm. It had that level of learning.

They didn't collect data on any of this for

human/machine interaction, which is a terrible waste. But it was ahead of its time in that sense.

In popular culture, you'll see documentaries of these fixers of Sony Aibo, and these people who usually -the ones that the television picks up at least -- are elderly and they are living in remote areas. They are so happy when their mechanical dog that they've had for 16 years is actually lively again. And they were so sad when it wasn't.

So now there's a range of industrial applications that are going into health care like a really, really, cute white seal that you can hold. And if you're bedridden, these things can -- they take a bunch of sensors, sensory input from the fur. These are portrayed as wonderful potential solutions to a massive, massive problem.

It's a little different than how it goes here. These are your friends. The idea of having sensor-laden machines in hospitals for various people is seen as a potentially great thing. So, again, money is pouring into this kind of area, both to the companies developing it, and to the researchers.

Transportation. Uber is not in Japan yet but the interesting part of transportation is after decades of pork-barrel politics, where the political strategy for the

dominant party that's been in power almost continuously since 1955, is to pour money into infrastructure projects, building bridges to nowhere or more like small villages with small numbers of people. So you have this infrastructure throughout Japan -- mountains, landslides, things like that.

The highways are great. There's nobody on the freeways. There are nice roads next to the freeways that are free, but there's nobody on either of them. There are nice bridges to tiny little hamlets. But the train lines, because the population is decreasing, are going under.

So buses are a strange growth area. High-end management consultants that now went into the bus business, are saying, Actually, it's local buses. And this is very profitable because everybody now, they don't drive anymore once they're older. They need these buses. There's huge demand.

What's the problem for expansion? It's the lack of bus drivers. So automation would be great. We have great infrastructure. We have a great need. And just operate them like trains, people are used to trains. And they have automated trains, or trains that have one driver that don't do much. And remember, we're in an area where there's high speed rail -- every three minutes, there's a high-speed

train that goes 280km/hr from Tokyo Station to different parts of Japan. They leave every three minutes, and with one of the biggest earthquakes that the world has ever seen, not a single train derailed. So people have great faith in technological solutions.

And the train drivers don't do much, right? We know that, and they know that. And that's fine.

So buses. This is great potential. And same thing, money pouring into R&D. Automated driving in Japan is not going to come from things like Uber, but it's going to come from this kind of regional transportation because the funding into regional revitalization gets you funding very easily. And so transportation ...

So all this funding, yes, well -- society is deeply wealthy. And the relatively equitable distribution of wealth, partly through these infrastructure projects and whatnot, means that there's a lot of wealth to tap into, even in the regional areas. And a lot of this has been through providing jobs by propping up small and medium businesses.

That's the social security model. A lot of these are zombie firms, and it's been analyzed that it's helped Japan not grow for the past two decades, but social stability was guaranteed. There was a tradeoff between social stability and economic growth, if you put it that way.

However, the main challenge to small/medium enterprises that have been used to prop up people is now the proprietors are retiring and these businesses are not replaced. That's fine, they were zombie businesses. But now those workers are not getting pensions from companies. So instead of universal basic income, Japan is facing a large proportion of the population that is depending on government social security anyway. It's almost like a universal income and it's shifting it out of companies.

So automation that drives away small/medium companies that are not productive is almost fine because there's a high attrition rate in them anyway. So that's why this fear doesn't pop up in political debates.

Replacing highly skilled people. The issue with highly skilled people being replaced by AI and robotics is not that the people are going to lose their jobs. The debate revolves around these people who have lots of skills other Japanese companies rely on, and a wide range of suppliers that are good at one or two things, and they keep moving into high-end -- well, the people that are doing those things that can - again, in the popular culture, they love these things. You have a machine controlled ball

bearing manufacturer. And then you have somebody that's really, really good at carving things, and they have a team -- they keep making improvements, and of course the highly skilled craftsman ball bearing goes about four times (that ?) the highest of the highest-end university researchers could produce.

This kind of skill isn't being replaced and replenished. So if you can capture this with robotics and AI, that's a good thing, because the people are disappearing that provide the underlying competitiveness to a lot of Japanese industry. That's how this is framed. So getting highly trained people to transfer their skills somewhere else before they disappear, this is seen as a good thing.

So within various ministries and levels of government, everybody wants to get on the bandwagon of AI and robotics, which means that each ministry actually at this point, they can bundle all sorts of things under this umbrella, and get government funding for it. And the government, the Abe administration -- Abenomics, as it's called -- prints lots of money, spends lots of money, but a lot of the money that is now a fiscal deficit is owed to domestic sources, so it's strangely stable.

So lots and lots of money comes down. There's

regular budgets and supplemental budgets. When the supplemental budgets hit, it's a couple of trillion dollars that need to be spent very quickly. So you have the Ministry of Finance middle level officials who suddenly have to dole out money in billion-dollar chunks. And so it's easy to spend these billion-dollar chunks in AI and robotic laboratories that are then affiliated to each ministry. So the Agriculture Ministry has one; the natural leader, the Ministry of Economy, Trade, and Industry, they have several. And then coordinating these, the cabinet office, and there's a jostling of who controls the cabinet office. Part of that linkage with Research Institute link came with basically a billion dollars over the next decade to do what they want and work with about 70 companies that I've signed on, and make an AI center right near Tokyo Station.

So there are lots and lots of initiatives going forward, and lots and lots of money that lands in these places. The obvious question next is -- what are the worthwhile things for this money to actually go? There will be some good things to come out of it; a lot will be wasted. But the logic is now, Get as much as we can now because it makes a lot of sense.

So what are a lot of the industrial deployments

that are interesting, because a lot of money can be wasted? And this is my last point, which is Komatsu is a Japanese construction vehicle, construction equipment company. They've had automated mining dump trucks, those huge ones that operate in mines, fully automated since 2008. But they make mining dump trucks and a lot of other construction equipment. They're not consumer facing. It's a relatively closed environment, so you haven't heard of them.

What they like to do now, and what they're doing and deploying is using diggers. So for example, it takes ten years of experience, apparently, to use a circular, one of these diggers, to carve a slope. And if you're going to build a big building, you can't dig a little too far and Play-Doh it back and make a skyscraper. You could, but they don't do that in this country or Japan.

So ten years of experience to do this. However, Komatsu's autopilot, can reportedly, it can basically be your first time operating this thing -- I think that's a bit of an exaggeration -- but you don't need ten years of experience. You just bring the device to where it needs to go and hit "autopilot" and it will stop you if you're about to make a mistake. And it will assist you in making these cuts.

So as a paradigm, this is very interesting,

because it is basically deployed. They've got a bunch of tests; they haven't scaled it, but they're talking about it. And the Ministry of Construction is going into bidding that requires this kind of thing. We're talking about lowskilled people performing previously only high-skilled people possible jobs.

So what's the range of jobs that disappear with AI? Oh, low-range, you have to be highly skilled. Well, here's actually a whole lot of high skilled jobs that can be performed by low-skilled people. What kinds of areas will this kind of development occur? Well we don't know yet. But if it can be done by Komatsu, which is not known for -- they were very fast to go make a deal with a drone company to get the data so their drones flying get realtime data to control the machines, so they're a little unusual, but still. What high end jobs there can lowskilled people do to upskill?

And then for the highly skilled people, what do they do next? In Japan's case, it's easy because the bottleneck for construction sites is that there aren't enough people with those ten years of experience. So that's why it's very easy to deploy this stuff.

That's a paradigm to look forward to. So all the numbers of these are the jobs that will be automated and

these that aren't -- what if we apply this paradigm. At a technical level, I don't know, I think it might be almost unknowable at this point. But there are people working on it. And what we see out of Japan is importantly industrial application of these kinds of things on a broad scale, with no political opposition or fear about taking people's jobs, but enthusiastic support to deploy this stuff.

And so that's where my current remarks will stop, and we can talk about some other things like why do they think like this? It has to do with lack of a military. Right, there aren't human-killing drones and things like this. But I'll end there for now. Thank you.

DR. ZYSMAN: Thank you, Kenji. I want to ask the speakers, and I think Dan is still on the line, one question, and then maybe take a couple of questions from the audience.

I think one of the questions is when we see innovation in the United States, over the last years, and particularly in IT, we have always assumed that our market would be the driver, and that in fact everywhere else would be where we would export our technologies. The question that becomes interesting is to what extent do these foreign experiences, and not just these, but others, represent not just lessons and challenges, but sources of competition and

markets in which we become the market to which others export rather than necessarily the other way around? What is the balance that's going to shift?

Much of the IT story until now has been a consumer-driven story. It's been a social media kind of story. And the stories we're hearing in all of these cases, even including the security side from Danny are very different in nature. So Danny, if you're there, maybe we'll start with you since you're off there somewhere up in the wilderness. Can you hear us?

DR. BREZNITZ: Yes.

DR. ZYSMAN: Do you want to respond to that? And then since we can't see you, we'll then turn to Kenji and Robin.

DR. BREZNITZ: Sure, up here in the North, other side of the wall, which you are soon going to build, I would agree with you. I would actually, again, looking at my research on the IT industry, John, you've always beaten the market. Actually, your importance in the market is actually being reduced, which I think would also mean that your ability -- I'm an American, our ability; I'm paying taxes in the US, I might as well say "our" -- our ability to come up with innovation, especially here, would be limited. I think Kenji's presentation made it very, very clear. Since a lot of that, I don't think, is going to be basic fundamental research, but actually the application of things we already more or less know how to do in ways which allow organizations to make money and/or provide services that are needed, what we just heard about Japan, with the buses, just raised my enthusiasm for those kinds of things. It might mean that Japan, or either Japanese companies, or companies operating in Japan, have a better lab, if you will, in order to find those solutions, perfect them, and then export them.

I literally do not know whether here in the US, we, for those kind of solutions, have two things that we used to have in the past: a) the best environment, what I call the laboratory, to try those things out; and b) because we're talking about services and about privacy and about data and about a lot of things that are heavily regulated, and it's not just consumer tastes, I am not sure that we even have the best environment in which to develop exportable products anymore. I think we do, but I expect that we'll find a lot more competition.

I also suspect that the extreme variety that we have in those kind of environments means that different experiments, if you will, will be played out in different

places. And some of our companies will be blindsided, which is absolutely fine. But again, if you don't prepare for that, we shouldn't be surprised.

Again, I don't hear you, if you want me to stop. I stopped. If you want me to say more --

DR. ZYSMAN: I'm turning to Robin now, Danny. I know you can continue if we let you.

DR. BREZNITZ: Always.

DR. MISHRA: Only short answers because I don't think we can predict the future.

Of course, there is competition. As I tried to make clear, there is a consumer business in the US, and Germany will not even try I think to compete with that. So the Germans are on Facebook, they have their iPhones, and then they are on social media, they may get all of the Amazon or Google devices and put them into their houses.

That's not our business model. We have a B2B model. And of course, it's under pressure, like the US model may be under pressure when it comes to implementation into the real industries, or in health care, or in education, because the system has to work, and then you can implement the artificial intelligence.

I would more stress the point of conditions for this competition for a situation where we all can benefit from that. My feeling would be that there are three gates that have to be kept open. So let me start with open access; I think there is a feeling that we have to share data, that it's not only about competition, but also about cooperation in certain spheres of society.

It's about open borders. So it's about trade, still. It's about exchange of products. I think it's still a crucial part, and we will see how the actual situation here plays out. I mean, there are trade-critical people in Germany as well. If you have industries who are really export-oriented, and I think Germany, US, Japan as well, it is still important to have this framework. And we are talking about open societies; our feeling is that it's really important that you have the framework of a society where your creativity can thrive and where you really have conditions to innovate.

Immigration is an interesting factor. You can have too much immigration, and you can have not enough immigration. But this is also a question that we have to ask ourselves as societies. Who will be the winner of this development. It's very difficult to see.

I think the Japanese model works well, because it's in a different section. Germany's model will always only work in international settings under certain

conditions. But my feeling is that the same is true for the US.

DR. ZYSMAN: Kenji?

DR. KUSHIDA: Yes, that is a really interesting set of points.

I think there's no doubt that many of the high end algorithms, like Deep Mind, which is originally from the UK, but empowered by Google, will come from the US. And of course, it's the deployment in different areas, or the ability to get different kinds of data that's not easily available here, where some of the things might be shaped.

Let's talk about Japan again. Many people have the image of the Japan as the 1980s, especially in Washington, where there was trade wars. It's about access, lack of access and all sorts of government and industry working together to actively block foreign entities from coming in. If it's packaged -- particularly if it's packaged as solving social problems, like the demographic crisis, then Japan now is almost completely open to foreign firms.

This is emphasized even by formerly very protectionist, ministries. They're saying, Look, we need help. Especially if it's consumer-facing, because there are no Japanese winners that are trying to block this. Foreign

firms come in and use the data that we can provide, actually tell us which data that we can provide so that we can work with people. Or things like the Trucking Association just came to Silicon Valley, Well, we're going to be disrupted. Let's figure out if we can work together by providing data that no one else can provide so easily, and let's be a first mover in that.

These kinds of things happen. And of course Japan is export-oriented, but only about 16 percent of the GDP is from exports the past couple of years. So even if there are health care systems and things like that, it's harder to export systems than goods and services. For the economy it's not obvious, but again, packaging these things as solving social problems, somewhere like Japan's deployment patterns might be a useful place to look at and see if something like that would work here. At the same time, getting data from there that you can't get out of China or out of other places might be interesting.

So there was a news article this morning. And because it was a news article, I don't know the details. But basically Toyota and 70 companies added ISO standards for managing information flows in factories. There have been movements around this. And I could see this happening where the big companies were trying to figure out how to

get good, next generation robotic systems in their factories, were looking to software algorithms and AI and things like this, but understanding that there needs to be a set of standards that they can work together for. And that Japan owning proprietary standards, every time they've done that in the past, it's not gone global. And so they're trying to get ISO.

And this is the part that I really don't understand, so I have to go figure this out. They claim to say that they want to make use of bottom-up knowledge flows, like the way Japanese manufacturing traditionally has, which is different from top/down optimization solutions that are then implemented elsewhere for what people are tasked to do.

So that kind of thing is also interesting. It may work out. It may not.

DR. ZYSMAN: Let us take one or two questions. As we do that, I would say that Kenji's last point here, having held discussions in a whole array of places, it's fairly clear that part of the basis of industrial competition is a fight about who's going to be able to shape the rules -- let's call them, broadly speaking, the rules -- of the game. And certainly Industry 4.0 is part of the German play to try and not allow that to be set in Silicon Valley.

Yes? Can you say who you are though?

MR. MACLEOD: Ian MacLeod, John Hopkins University Applied Physics Lab.

Professor Kushida, you finished your talk with a nod to the military. And that's actually my question about Japan. The self-defense force is actually much larger than people think. The demographic trends are not really in its favor for recruiting and retention. It does have very serious regional security concerns.

So why then, do we not necessarily hear more about adoption and application of autonomous systems and robotics in the Japanese military, for military purposes? Or is that not true; it actually is, we don't hear as much about it? Thank you.

DR. KUSHIDA: That's a great question. I think that there are developments underway, but nobody needs to shine a flashlight on it, and because of Article 9 of the Constitution, written by the US.

If you noticed, the Prime Minister was awfully quick to go meet the President-Elect. And now he's going to Pearl Harbor a little later this month, which is fascinating, because no one's ever done that, which is clearly a signaling mechanics saying, Hey, make sure your nuclear umbrella covers us for the coming years because we do have these conflicts.

And so if the immediate security threat is off the table, then they can do all sorts of things without entering the public eye at all. And it doesn't enter in the political debates. Actually, I do think there's a fair amount of money going into that, but they don't have to make it so clear.

DR. ZYSMAN: Even 25 years ago, or 20 years ago, it was clear that the Japanese discussion was, Why did you want to build a commercial satellite launcher, just so you could put ... as was said in a meeting that people did not realize there were outsiders, that you could put other things on top of the satellite launchers?

So I think this issue of dual use technologies is in fact a way around that conversation. And if it was ever necessary, it would become very rapidly visible.

Other questions, Kenji do you want to add one thing?

DR. KUSHIDA: By the way, Canon, the camera maker and precision equipment, they just entered the microsatellite market. There can be other things on top of satellites, other satellites that can go on that. So some of this is ... MR. NEW: Josh New at the Center for Data Innovation.

So over the past year or two, we've seen a lot of efforts in the US, and I guess in the UK, to make sure that a lot of basic research and training data, and building blocks of AI are democratized. So the most notable example is OpenAI. We have the partnership to ensure AI benefits. And there's a lot of momentum, I guess, in the academic community to have this really robust research sharing network, so that companies can of course invest in their own proprietary developments on top of these basic building blocks, but everyone gets that initial starting point to work with.

Are those efforts being replicated in all of your respective countries that you're talking about? Or is this primarily like a Silicon Valley-driven thing?

DR. MISHRA: Yes, I might pick this up. In Germany, we have really new rules for open access in academia. But what is interesting is that the publishing houses really tried to fight this, and still are working on their rights to keep the publication for themselves for about a year, and then open it to the public. I think the general direction is clearly in favor of open access and in sharing training data.

From my point of view as a science and an innovation guy, I can only support this. But there is of course, I think you have this same discussion in the US, there are of course companies who see it as their business model to be the first to publish things and to keep them secret for a certain while, to make money with it. Of course, these interests have to be balanced. There has to be some payments, or some structures that really allow these people who collect knowledge to survive. So this is the balance that we have to strike. But there are open access laws, and they really mean open access.

DR. ZYSMAN: Kenji, do you want to answer?

DR. KUSHIDA: Yes, so in Japan, we have this bifurcation where the research institutes and universities have lots of government funding in it, and to get more government funding, they work with multiple companies. When this happens, it's shared, it's pretty competitive, this kind of thing.

The things that have developed within companies, and then bringing academic input into that, like a big manufacturer, robotic arm with the AI that can do pattern recognition and things like this, all of that is locked in and proprietary, and that's a competitive thing.

One interesting twist is that a lot of the

researchers tend to be focused -- and this is a critique from the outside, and I don't know the actual numbers, but there does seem to be an awful lot of focus on bipedal walking robots. And for those startups like from Tokyo University, that got bought by Google, everybody wants to make walking bipedal robots. This goes back to norms and back to military a bit too, but robots are your friend, and this is actually a deep, deep-seated norm, and humanoid robots in particular.

And so a very strange set of discussions often goes into very quickly -- robots, okay. We have the industrial robots, and we have the human friend walking robots, but not the killer robots that walk around.

So short answer: structurally, a lot of the funding goes into things that are forced to be open because there are many companies; another set of things that are being done in industry are not open at all. But it's not obvious they're cutting edge, so that's what the researchers that are in the field can tell us. And then this focus on things that are encapsulated in mechanical, and all of that is closed because it is proprietary.

DR. ZYSMAN: Thank you very much.

(Break)

MR. GREELY: Good afternoon. I hope everyone is
ready for our closing keynote address of today. It's my great pleasure to introduce Arati Prabhakar, who has been, since 2012, the director of DARPA.

What I find particularly interesting in her biography is a couple of other things. One is she was on the board of STEP, one of the two organizations, along with the Committee on Science, Technology, and Law, which has been co-hosting this event. But I think more importantly, like my wife, she grew up in a small town in a remote part of Texas, and somehow has managed to -- can I say escape? -- is no longer there, but also along the way spent 15 years working in Silicon Valley.

So I'm going to claim you as one of the many Californians at this meeting, even though you've been in exile for the last six years or so, and welcome you to take the stage.





Presentation of Arati Prabhakar, 12/12/2016

DR. PRABHAKAR: Thank you, Hank. And how great to be here with all of you. It looks like you've had a phenomenal day. You've had all kinds of perspectives from some really amazing people. We'll find out if there's anything left to add. Go like this if you've already seen it before, and I'll just zip on through.

But what I wanted to do was to share with your our perspective from the lens at DARPA on where artificial intelligence is today. I think most of you know, but DARPA is part of the Defense Department that was created in the wake of Sputnik to prevent that kind of technological surprise, and we pursue that mission by creating surprises of our own.

We focus on military capabilities for national security, but in doing that work, we also over many decades now have invested in the core underlying technologies that with a vast amount of private investment afterwards, have really led to huge changes in how we live and work as well as in national security capabilities.

And in fact, artificial intelligence is one example of exactly that. When we think about artificial intelligence in the DARPA context, we think about two things. One is how do we use the advanced technology that exists today? And the second is, how do we advance the technology for the next generation?

Powerful, but limited is very much the way we think about AI today. You've been talking today, and every time you open the paper, you see another powerful example about what artificial intelligence can do. We think there are very important applications for national security, and I'll show you some of those.

But at the same time, we think what we're seeing in AI today also has some quite fundamental limitations. And those have to inform the applications that we pursue, but they also point the way for the next research agenda. And I'll give you some examples for some of the new things that we're pursuing as well.



Artificial intelligence

Three waves of AI technology (so far)

Handcrafted knowledge Statistical learning

Contextual adaptation

The first wave is still advancing and solving hard problems The second wave is amazingly effective, but it has fundamental limitations New research is shaping the third wave

Presentation of Arati Prabhakar, 12/12/2016

I think the place to start, with the first slide please, is I wanted to share with you the lens through which we see the technologies of artificial intelligence. I had this marvelous moment; about a year ago I was on a trip with Secretary of Defense Ash Carter, and one of the group of staff members was trying to tell him about what was going on with an AI initiative. And Secretary Carter wrinkled his brow and said, Isn't artificial intelligence a really old term? Which I thought was marvelous, because of

course it is a very old term.

And what I think is happening today, and the reason for this enormous focus on AI is the fact that a second wave is now breaking over us. And I'm going to walk through how we see these three waves of artificial intelligence. The first one, about handcrafted knowledge expert systems -- my key message about that is that it's old, but it's not boring. It's still advancing, it's solving really interesting new classes of hard problems.

But today a lot of what is happening is the second wave of statistical learning. And these new class of AI systems are very, very effective for a really different, much broader class of problems. And this is where we see a lot of the new power of AI technologies. And I'll give you examples of that, but yet it still has some quite fundamental limitations.

And those limitations, of course, then point us to some of the ideas that are still forming about what the subsequent generations of AI technology might look like, and how new research can shape it. Whatever happens, in retrospect we'll call it a third wave, but it's still amorphous and multi-faceted. Okay, if we could go to the next slide.



Intelligence is an ability to process information

rich, complex and subtle information

learn within an environment

abstract to create new meanings

reason to plan and to decide



Artificial intelligence is a programmed ability to process information

Presentation of Arati Prabhakar, 12/12/2016

Just to say a word about how we're thinking about the term intelligence. I think people like to say is the system intelligent or is it not. We think of it less as an on/off switch and more as degrees or intensities of capability. I think it's useful to think about different dimensions of systems' or humans' ability to process information, whether it's perception, learning, abstraction, or reasoning. I'll use this little bar to indicate different degrees of capability as we talk about artificial intelligence, machine systems' ability to do these different dimensions of intelligence. Why don't we go to the next slide?



First-wave AI technologies



Presentation of Arati Prabhakar, 12/12/2016

So we'll start with the first wave of AI technology, this of course dates back many, many decades now. And really, the idea here was to encode expert knowledge and to create a series of if/then statements, if you like, in well-defined domains -- that's the part that engineers did. And what came out of that was AI systems that could reason over narrowly defined problems. Many, many examples. Planning tools in my world, the Command Post of the Future. If you used a system like TurboTax, it converted the tax code into a series of if/then statements that you could then, it would become a very able assistant for complex tasks, like filing your taxes.

And cybersecurity. Software analysis is an area that I think is a particularly good example of where this first wave of AI systems is still making some very powerful contributions today. A great example is a recent program at DARPA, the Cyber Grand Challenge. This is one of our big prize competitions that we held. In this competition, we challenged teams to build systems, machines, that would do the kind of reasoning that today humans do about code as cybersecurity professionals, or as hackers, look through code and look for vulnerabilities, as they design patches, and they look for opportunities, if they're hackers, to exploit flaws in others' code. All of those are processes that happen with intense human effort, often can take months if not years on sophisticated code. And what we demonstrated this summer in the Cyber Grand Challenge is that indeed, largely through first-wave AI technologies' expert systems, it encoded that kind of knowledge and expert capability into machines.

We were able to demonstrate machines that could identify flaws in software, generate patches, respond in timescales of minutes rather than months. Dramatic advance in capability, and one example, I think, of some of the kinds of things that are going to lead to a future where we can win at cyber, something that I think we all understand

is necessary. In fact, if you're going to build AI systems that are critically reliant of having software and data that you can trust, that's a whole story in its own right, about cybersecurity, but a great example of where first generation, first-wave AI, I think is still advancing.

Now, some very significant limitations to this generation of the technology. No ability to learn; no ability to handle uncertainty. Ironically, we no longer call things like OCRs (optical character recognition) artificial intelligence, because now we just take it for granted. But remember how dazzling it was when machines could first do these things. That's what that first wave was really -- for those of us who remember experiencing it, that's what it was about. So let's move forward.



2004

completed: 0

First wave stumbles on natural data

DARPA Autonomous Vehicle Grand Challenge 140 miles of dirttracks in California and Nevada



2005 # completed: 5

The problem in 2004 Vehicles were able to follow the GPS waypoints very accurately but either missed or hallucinated obstacles abead The difference in 2005

A probabilistic algorithm interpreted the flood of incoming sensor data to learn the terrain and map out an optimal driving surface

Presentation of Arati Prabhakar, 12/12/2016

One of the places where first wave really broke down, though, was in dealing with natural data. Data that was statistical in nature, where understanding and recognizing patterns that have been presented in different ways, where that was a kind of problem. A great way to think about the before and after picture of second wave AI technologies is what happened in a different DARPA Grand Challenge.

About a decade ago, we started running prize challenges that challenged people to build self-driving cars that could autonomously complete the course. Quite famously, the first one we did in 2004 was so tough that no machine was able to complete the course. And part of the problem that was going on at that period was that while these machines could follow a GPS -- they had GPS, they could navigate to way points -- but they weren't able to deal with the obstacles in the way, and sometimes they made up obstacles that weren't there. So quite fundamentally limited.

A year later, we ran a similar challenge again, and that year, five contestants were actually able to complete the course. One of the major advances that happened in the intervening period was the use of probabilistic algorithms to start being able to take that flood of data, and turn it into an optimal driving path. I think this was a great, a little glimpse, a decade ago, of this generation that was about to come. So let's go to the next slide.

DARPA Second-wave AI technologies					
	Handcrafted knowledge	Statistical learning		(8)	Apple Siri
	Engineers create sets of rules	Engineers create statistical models for specific problem		Y	
	well defined domains	domains and train them on big data		Google	Text analysis
	Al systems reason over narrowly defined problems	Al systems have nuanced classification and prediction capabilities			Image
	No learning capability and poor handling of uncertainty	No contextual capability and minimal reasoning ability		Fairface, Sabardan, and L12015	recognition
	Perceiving Learning Abstracting Reasoning	Perceiving Learning Abstracting Reasoning		and the second	AlphaGo

Presentation of Arati Prabhakar, 12/12/2016

And today of course, this is the era of the second wave of artificial intelligence technologies. It's an era of statistical learning. And now, in the systems that we're talking about now when we talk about this generation of AI are systems in which engineers have created statistical models. They are working in specific problem domains. And then with the ability to train those models, this is how we're now able to build artificial intelligence systems that can do these amazingly sophisticated classification tasks, a whole new set of capabilities.

And the examples, I think, are pervasive. In our consumer lives, we see them in what's happened with natural

language processing, what's happened with search and text analysis, what's happened with image recognition. You go on Facebook, and a picture you didn't even know someone took has automatically been identified with an algorithm as having your face in it. It's that sort of amazing moment where all of that seems very exciting and fresh and new.

And for those aficionados who followed Deep Mind's amazing win when their system AlphaGo, an AI system that plays Go, competed against the world's leading Go player. That was an amazing moment of watching a machine be able to do something that was a task well beyond the capability of this first wave of AI systems. The number of moves in Go is so vast, that you simply can't compute all of them and make your choices.

And so interestingly, AlphaGo is I think a cool example of a system that used what it had learned from playing lots of games to build a tree of possible moves, to limit the search space, and then use first wave techniques to map out which of those might be the best place to go, and I think an interesting glimpse into the power of starting to combine these techniques. An enormous amount of capability, but again, some important limitations.

If we go to the next slide --



Presentation of Arati Prabhakar, 12/12/2016

You know, it feels like magic when you start to see these technologies. It is not magic. These advances came from the confluence of many different kinds of advances. And one of them had to do with the machine-learning algorithms that had been developed. There was a huge surge in capability with deep learning systems, big advances in the neural net architectures that had been used. That's only one of many advances in the learning algorithms that are being employed.

As I mentioned, these systems learn on vast amounts of data. And of course, in just these last few years, we've continued to have this explosion in online data. Think about how many images are now available to train your systems on. That's been a big factor.

As well in the hardware world, when people are trying to build these kinds of deep learning systems using CPUs, they were spending many, many millions of dollars for big systems in a warehouse. The GPU revolution, with the chips that were designed to do graphics processing, also had a marvelous architecture for solving this class of problems. That crashed the amount of processing power and the cost that it takes to do these kinds of problems. So now you can do interesting things with the amount of computing power that's on a desktop. That's been a very big advance.

And then finally, of course, as these deep learning technologies, as the second wave of statistical learning, is surging forth, there are so many applications pulling this technology forward. This is a map of where patent activity has been over a period of time. But I think it gives you a sense of a very great variety of commercial applications that have become an enormous driver. We can go forward from here.

Just to give you a little bit of a sense of how we see applications for this second wave of AI in the national security context. We do see enormous new

possibilities. One example is -- a decade ago, we were doing the self-driving car challenges; now the commercial industry is off driving that field.



Today's artificial intelligence is powerful...



Operate highly autonomous platforms

Presentation of Arati Prabhakar, 12/12/2016

We've been off building a self-driving ship. This spring, we christened the Sea Hunter; it's a ship that is able to leave the pier and navigate across the open oceans without a single sailor on board. It has sparse supervisory control, a sailor watching it from the United States. It's not a remote-control ship. It's able to function at a highly autonomous state but still have that human supervisory control.

In order for this ship to be able to do the kind of complex navigation that it needs to do, it needs both first wave and second wave AI technologies, both to follow expert rules about navigation but also to do the image recognition that's necessary, especially in congested conditions. So a great example of an application in my world, and if you can click again --



Today's artificial intelligence is powerful...



Search the deep web

Presentation of Arati Prabhakar, 12/12/2016

Another example, again, an analogy to something that is happening in the commercial sector. One of my program managers a few years ago went to a conference on human trafficking. He came back, heartbroken, really, and enraged to find that in human trafficking circles, it's possible to buy a human being for \$90. This enraged Chris so much that he decided to go do something about it.

And when he spoke with our partners in law enforcement, what he found was that they search the web looking for the signs of human trafficking networks. They were limited to the same kind of search tools we have. So number one, it's a single threaded search through a vast space; but number two, commercial search engines of course are only indexing the surface web, the small fraction of the online available information that is commercially viable and useful for them to index. And the rest of the web, the deep web, and the dark web, are of course wonderful places for this kind of illicit activity to hide.

So in this program, our Memex program, one of the things we did was build tools that search the deep web and the dark web, and again use many of the techniques of AI, for example, to identify relationships among images or analyze similarities in the way code is written. Again, look for patterns across that vastness so that we can create maps that they can point our law enforcement partners to the richest directions to search to find human trafficking networks.

This tool is now being used by DA offices across the country, and to date it has led to -- it's been used in hundreds of investigations, and those investigations have led so far to 27 indictments and three convictions, and I hope there's much more ahead. It gives you a sense of one example of the kind of powerful tools that could start to keep up with the pace at which online information is



Today's artificial intelligence is powerful...



Manage cybersecurity in real time

Presentation of Arati Prabhakar, 12/12/2016

A very different dimension of cybersecurity, another example of where we're using AI techniques -- this is a program called Network Defense. It looks at the tens or hundreds of billions of individual network transactions in a complex network, for example, like the one at the Defense Department operates. And in that vast, vast, vast quantity of network transactions, it's able, using some very sophisticated math and techniques of data analytics and AI, to identify really tiny anomalies, just a handful of network transactions that might be an alert, a low false alarm kind of alert, an intelligent alert for an analyst or a cybersecurity professional, a very different dimension of cybersecurity but another one I think can be very helpful.

And then finally, the last example that I'll share with you is about spectrum. All of these examples are AI tools to deal with data of different sorts.



Today's artificial intelligence is powerful...



Overcome spectrum scarcity

Presentation of Arati Prabhakar, 12/12/2016

Spectrum allocation is a data-rich environment that I think offers a completely different kind of classic problem to tackle.

We just recently launched what will become DARPA's next big prize challenge. This wasn't going to be a talk about prize challenges, but somehow I've touched on a bunch of them. This new one is called the Spectrum Collaboration Challenge. And here, we're challenging teams to build radio networks that embedded artificial intelligence. What those AI systems will do is scour the spectrum, and form hypotheses about how others are trying to use that slice of spectrum to communicate, and then based on that, it will form its own pattern of how it wishes to try to transmit its data.

And it's going to be a very unusual challenge in that the competitors will get points for transmitting their data, but they'll also get half points for what the other participants are able to transmit, because what we're trying to do is build systems that can collaborate to maximize the amount of data that can be crammed into a fixed amount of spectrum. If you think about how many conflicts we already have between defense needs for spectrum and commercial needs, and then you think about the where the Internet of Things is going to drive wireless data demand, you start to see how important it's going to be to maximize. And through some of these techniques, we aim to achieve many multiples of the amount of wireless data in a fixed amount of spectrum. I think it's a very different example of the kinds of things that are possible with these kinds of adaptive learning systems.

Okay, super powerful, lots of examples of places

where we are finding huge advantages based on AI. I got curious and so a couple of months ago, I asked people to look across our entire portfolio of about 200 programs at DARPA, and asked where were we using first wave or second wave AI. And while definitions are a little fuzzy, it's sort of startling to see how many places it's showing up. It is such a useful technology area right now. So let's move forward.

There's a lot to be said about how powerful it is. Let's take a minute and talk about where it's limited. These kinds of statistical learning systems -- it seems that every week you read about yet another statistical learning system that has exceeded human performance. In one area as an example is machines that can automatically identify what's happening in an image. Here's an example of one where a machine is able to label an image like this.



...but limited



Construction worker in orange safety vest is working on road

Andrej Karpathy, Li Fei-Fei

Presentation of Arati Prabhakar, 12/12/2016

Statistically better than humans. That's pretty amazing. It doesn't mean perfect.

And the problem is, if you'll click -- when these machines make mistakes, they make the kinds of mistakes that no human would ever make.

DARPA

...but limited



Construction worker in orange safety vest is working on road



A young boy is holding a baseball bat

Andrej Karpathy, Li Fei-Fei

Presentation of Arati Prabhakar, 12/12/2016

This is a hugely important thing to understand. And one of the things that we spend a lot of time doing at DARPA is being very clear with the potential users of AI technologies, about these kinds of limitations.

So when I see this, I think this is a really important message that if you're using an AI system to look at something that's statistical in nature and where every individual decision has low consequence, this is a great way to go. It's a great way to look at bits, wireless data being transmitted in the spectrum. It's a great way to look at micro-trade transactions and financial data. It's fine for Facebook. If they misidentify my picture, it's not the end of the world.

But if you are going to try to use these systems for high-consequence decisions, whether it's for a selfdriving car, or in my world, for targeting, this is an important flag that these technologies are not ready today. And we don't know if and when they will be ready to be able to be trusted with those kinds of extremely high consequence, life or death decisions.

This is one example, if we go to the next slide, here's another example.





Internet trolls cause the AI bot, Tay, to act offensively

Presentation of Arati Prabhakar, 12/12/2016

This is something -- did this come up earlier today about Tay? Tay was a chatbot created at Microsoft Research. It seemed to generate all kinds of perfectly reasonable humanlike responses when people played with it at Microsoft Research.

Then they took it out into the world. And I stole this slide from John Launchbury, one of my office directors at DARPA. And he likes to say that this was the least offensive thing that he could find that Tay said. We've all had the experience of what human behavior is like in the online environment. I think this is a great example of showing how AI, this learning dimension of this generation of artificial intelligence, in some really fundamental way it is an amplifier of what it learns and what it is learning from is an expression of human behavior. We know how awesome humans can be; we also know about this dimension of humans.

Now this is critical, because again it's about where we trust these systems, how they interact, how they reflect bias. But again, in my world, it's also an important flag that reminds us that a system that is trained is one that has yet another way to be messed with by an adversary. And in fact, people are already trying to think of ways to change how AI systems work by changing the training to which they're subjected.

Okay. So what do we do about these kinds of problems? Number one, we need to understand what that

means, about how we use the technologies today. But at DARPA when we see these kinds of fundamental limitations, it also points us to the new directions that we might want to pursue for these research-oriented programs. And when you go to the next slide, I'll show you an example.

DARPA Future	ologies	
Handcrafted knowledge	Statistical learning	Contextual adaptation
Engineers create sets of rules to represent knowledge in well defined domains	Engineers create statistical models for specific problem domains and train them on big data	Engineers create systems that construct explanatory models for classes of real- world phenomena
Al systems reason over narrowly defined problems	Al systems have nuanced classification and prediction capabilities	AI systems learn and reason as they encounter new tasks and situations
No learning capability and poor handling of uncertainty	No contextual capability and minimal reasoning ability	Natural communication among machines and people
Perceiving Learning Abstracting Reasoning	Perceiving Learning Abstracting Reasoning	Perceiving Learning Abstracting Reasoning

Presentation of Arati Prabhakar, 12/12/2016

Oh. Let me start by just saying a word about what this third wave might look like, as we pursue these research directions. And I would tell you this last piece about contextual adaptation is very much an aspiration today, and we're just starting to build out the portfolio of things that we at DARPA believe will be the research directions that advance, that there's much more going on. We're working with universities and companies.

But many different people are seeing this next generation in many different ways. I think that it's a particularly interesting time. But when we think about it, we think about a generation in which engineers can build systems that themselves are able to start building these more explanatory models. And again, to be able to work across many classes of real world phenomena. We're interested in exploring whether AI systems can actually learn and reason not in one specific domain only, but take what they've learned from one area, and move that contextual understanding into a different arena. And we're starting to imagine how things will advance, as we start fundamentally changing the nature of communications between machines and people.

If you go to the next slide, I'll give you a couple of examples of some of the programs.



Presentation of Arati Prabhakar, 12/12/2016

One is a new program called Explainable AI. Today when a deep learning system gives us one of these head scratching results that makes absolutely no sense, you take the cover off the black box, and what you see is a bunch of weights. Then what do you do? How the machine arrived at that set of weights from the millions of images on which it was trained, is a complete mystery, and it really doesn't leave you with a next move.

The objective in the Explainable AI program is to develop systems that maintain that very high statistical performance, but that build in explainable models and an interface that would explain to a human user where the elements of the answer came from. So instead of simply saying that this image is a cat, this future system -- our hope would be -- would explain that it arrived at that decision because it saw whiskers and claws and that it had these cat-like ear features. These kinds of advances in explainable AI are going to be critical if we're going to start being able to trust AI systems, and then advance them and use them for a wider set of applications.

Another example, on the next slide ... virtually everything that is going today in the world of big data is about finding correlations.



Early result: An automatically generated web of influences from 1000 research papers Presentation of Arati Prabhakar, 12/12/2016

My Silicon Valley friends like to say if you have more data, it's always better. And I say back to them, if you have more data that somewhere in the data is the data to support your hypothesis however deeply flawed it might be. And I think that this is a major concern in a world in which correlations are all we're doing.

And commercially, correlations are so valuable. When you do AB testing you put two different things out on the online environment. If all I want to know is what causes people to click more or buy more, I can close that loop very, very rapidly. I can do more of A if that's what works, and I can keep iterating and experimenting. That is so fruitful that I think that we're going to go down that path for a very long time.

But all the really hard interesting problems are not just about correlation. We're going to really have to start developing our ability to understand causality. Today, that kind of reasoning about the fundamental linkages and the cause and effect relationships, that is reasoning that is left to human beings. We've done pretty well. We've advanced the state of science for many, many generations, and many eons.

But we are still wrestling in problem areas where there are many, many different factors and they feed on each other, and they loop and they iterate. In these complex environments, we still struggle to get our heads around the causal relationships and how they're working together. That's what we're going after at DARPA in our Big Mechanism program. It aims to be able to build machines that can knit together maps of causal relationships, and build whole system causal models, including quantitative models underpinning them, by automatically ingesting up to millions of scientific observations.

This particular program started by looking at cancer research -- that's an area where there's a lot of research that can be ingested using very sophisticated natural language processing techniques, and new semantic models. These systems are starting now to be able to assemble very sophisticated complex models of causality and causal behavior.



Uncover causal relationships among 1,000,000s of scientific observations



Early result: An automatically generated web of influences from 1000 research papers Presentation of Arati Prabhakar, 12/12/2016

This graph is actually an example of an early result from that program, with some folks at the Gates Foundation who were interested in the linkages between obesity and childhood malnutrition in particular. They sort of threw a bunch of papers, some research papers at our tool to see what it would do with them. And very quickly, it was able to generate this kind of a model, which I think starts to give you a sense of the kinds of maps and models that I think will be very powerful. And as humans and machines collaborate in this deeper way, I think we'll start opening new dimensions of problems that we think of as impossible today, I think over time will become increasingly tractable.

So let's go one more.

DARPA

Some third-wave AI technologies

Human-machine symbiosis

Automatic whole-system causal models

Continuous learning

Embedded machine learning

Explainable AI

Presentation of Arati Prabhakar, 12/12/2016

So those are a couple of examples. I mentioned explainable AI. A couple of the other things that we're thinking about in terms of what might be part of this third wave. Today machine learning takes place in the cloud, and that's awesome if you have a cloud at your disposal.

For much of what we need in defense applications, we need to be able to do it without access to the cloud. And so we're very interested in building hardware, special processing architecture. I showed you the leap of capability as we went from CPU to GPU. We think that's only the beginning, and so we're working today on some advanced architectures that again will be important for embedded machine learning in our world and will have much bigger implications as well.

We're pursuing a new program on continuous learning today. Today these statistical systems learn on data, but by and large once they have learned on that data, they are then fixed and whatever happens in their future application, their weights remain fixed. And we think that there's some very interesting opportunities for continuous learning.

I mentioned automatic whole system causal models. And I just briefly touched on human-machine symbiosis. But this was an area that I was particularly interested in. One

of our more famous predecessors at DARPA was J.C.R. Licklider. If any of you have read what he wrote in 1960 in a paper about "man-computer symbiosis", he talked about a future where humans and machines would do things together that neither had ever been able to do before. He talked about thinking as no human brain had ever thought before.

And one of the things I loved to reflect on is that was 1960. Think about how far all of our technologies have come in all these decades that have passed, amazing things that have happened, partly because of the inspirational visionaries like Licklider. And yet that vision about humans and machines teaming up together, that is still a vision that is out ahead of us. I think that is still for us to try to reach for.

While a lot of the discussion about artificial intelligence is about what happens on the machine side, I think actually some of the most interesting and most important advances will come as we fuse together how humans and machines tackle problems in ways that open horizons that neither of us are going to be able to do on our own. One more slide, please.



DARPA R&D investments in foundational technology and applications

Presentation of Arati Prabhakar, 12/12/2016

Let me just finish by saying that I hope I've given you a little bit of a sense of the things that we're doing at DARPA. Some applications of current technology, foundational research that we hope will be part of shaping the next generation of AI. If you take a cut through a particular area -- this one is natural language processing -- that's a long history at DARPA, of making these kinds of investments, both in the technology and the applications. And there's a long river that flows through the work that we've done, commercial applications of various sorts, as well as defense and military applications.
And I think this is a great snapshot. Welcome to my world. This is what we get to do. It's awesome and it's great.

I hope that's helpful in the context of the things that you're weighing. I know you've been talking about many, many, many of the other social implications and ethical implications of AI. Those are issues that are very much on our minds, and I hope that I've been able to share with you some of the dimensions in the national security context.

I would be very pleased to take your questions. (Applause)

MR. GREELY: We have about 20 minutes for questions.

DR. PRABHAKAR: Please, go ahead.

AUDIENCE: Thank you. That was wonderful. Other than throwing more money at the problem, what are things the next administration can do to accelerate the development of the current wave?

DR. PRABHAKAR: You know the current wave -- there is so much commercial activity going on because the applications are palpable and valuable. And so I think from where I sit in the national security context, I think it's important to be able to watch that and harvest it and integrate it with DoD secret sauce to work on the applications that are important for our needs.

I am sure there is much more to be done, still, in the second wave. To me, the research questions that are most interesting really address, as I mentioned, the limitations of what's going on today. To me, it always boils down to applications and then looking for what that next research agenda is. I think that's going to continue to be very fruitful.

AUDIENCE: One of the things we have been talking about are of course some of the ethical concerns related to AI. And just looking at autonomous weapons alone as an example. There's tremendous pressure from the military to increase capability, and the message is, We'll deal with the autonomy problem as it emerges.

But that's not totally reassuring. And I would hope that DARPA is beginning to look at what I would call the control problem. As these machines become more autonomous, these third wave capabilities, that you've built in some internal checks that don't entirely require human control, which may not be reliable down the road. Is that happening?

DR. PRABHAKAR: I am not sure where you are hearing the "We will deal with it later" problem because in

fact I think the department has been very clear about how important it is -- look, first and foremost, war fighting is a human enterprise. It's not the beautiful thing humans do, but it is a human activity.

One thing that I have really been struck by in my time at DARPA working with military folks -- when I talk to my colleagues in uniform, they are the last people who want to lose human control over what the machines do. These are the folks who are on the line. And so I will tell you, absolutely they demand advanced capability because we are in a competitive race. We're not the only ones who play in advanced technology in this era. So that driver is always there.

But at the end of the day, what I think the military is pounding the table for is capability so that it can be controlled and that can execute to what the human war fighter is trying to achieve. That's very much how I see it.

Now your question about how you build that into the technologies, I think, is absolutely key. And a discussion we have at DARPA all the time is about -- I think this actually starts with how you frame the way your approach research. If you frame it, and I talked to a lot of AI researchers who think of their grand calling as

making machines as powerful as they possibly can be.

I think that's a choice. It's not actually the choice that we have made. What we are focused on is how do we build human and machine teams and systems that can do things that you can't do separately. It's actually a very fundamental shift in outlook, if you think about teaming humans with machines versus just going in this one dimension and just thinking about machine capability. But at the end of the day, I think that's number one, more powerful, and number two, more tied to what our ultimate missions really are.

MR. GREELY: Way in the back.

AUDIENCE: The boat is terrific. With all due respect to the boat, and the planes, and the automated tanks, one of the terrific applications we have for AI is command and control. Once that gets implemented to a larger extent, how do you see that changing the nature of the relationship or the interaction that we have with our allies to the extent that we need to be worried about trust in these systems?

DR. PRABHAKAR: Can you say a few more words about what you mean when you think about AI for command and control?

AUDIENCE: Some of the weapons systems that you

describe that you're developing will be changing at such a rapid rate that the command structure of the military will be going into battle with multiple generations of weapons, with different squads having weapons that were five years older, two years older, or three months old.

They may not understand that they'll use AI tools to be able to assist in decision making. But that presents a trust issue, both for them, and more importantly, I would say for our allies to the extent that you would think about those trust relationships.

DR. PRABHAKAR: I think the picture your description is evoking is one in which there is a complexity inherent in the fact that you're dealing with multiple generations of systems, and trying to adapt to different capabilities that different people have.

That strikes me as something that throwing computing power at it, or throwing expert systems at it could actually be very helpful. I'm not sure if it's of the class of problems that lends itself to the kind of statistical analysis that we were talking about. And if it's a matter of simply using a machine to keep more accurate and complete track of who's got what and what kind of capabilities they have, I would tell you that that's a place where I think -- I'm not sure that has to do with AI, but I think it's a place where machines can actually be extremely helpful. I think they have been for many of those classes of problems.

I think the trust issue -- to me, the big trust issue comes when you're talking about learning systems. The power of their learning capability is also where I think it's important to be clear about where embedded biases are, and where they lack the kind of human context to make decisions in some corner cases that would be sensible from our frame.

MR. WALES: Mark Wales, from Adobe. One of the things that I've seen in a couple of DARPA videos and related is the idea of composable machines or swarm type autonomous things that combine for different purposes and so forth. Do you see a lot of potential for that in the DoD environment? Obviously, the ship is not quite single purpose but is dedicated purpose. Are you seeing a lot of potential both on the software and hardware sides, or that kind of composability and the innate generic purpose that could derive from being able to combine/recombine into different shapes?

DR. PRABHAKAR: Yes. I absolutely think that composable systems are going to be key to meeting the challenges that we have ahead, first because of the

flexibility and the kinds of architectures you can create and go buy, but then also so that in conflict, you are able to put together the assembly of components that you need to do the job at hand.

And I think this is a great example of a place where humans and machines teaming up to compose those systems is going to lead to a far better result than any other way of doing it. Number one, the complexity is too great for an individual or even a collection of humans to manage on their own.

I think the world is already full of things that we rely on computers to help us to keep track of and organize. But I think this is a great example of a place where very much the case where the commander in the field is the one who needs to be able to make a decision about what the objective is, but then wouldn't it be great if a machine put together a proposal for the combination of elements that would allow for the achievement of that objective, and then allowed the human team to iterate with its plan and achieve something that is optimized for the work at hand.

MR. GREELY: Other questions?

Well, that was a terrific talk. It gave us a lot to, I think, pull things together from the entire day.

DR. PRABHAKAR: Thank you. MR. GREELY: Thank you so much. (Applause)

Agenda Item: Final Remarks

MR. GREELY: Although I don't resemble him, and I think he might resent it, I am supposed to be Ryan Calo. Ryan, the organizer, the chair of the committee that organized this, is on a plane on his way back to Seattle, and somehow I have been asked to do the summing up. And the good news is I will be brief.

I think this has been a terrific day. Remember, I said at the beginning, this is not my world. I am a carbonbased life form, and I work on carbon-based life forms. So for me, this has been very new. And I am feeling bewitched, bothered, and bewildered, but that's actually really good. I am fascinated by this, and I think that I have a better understanding, and I hope the rest of you have a better understanding -- maybe not a better understanding, but a better understanding of what I don't know, and where the uncertainties are about where things are going.

That's the only substantive thing I wanted to say, except the most important thing of all, which is thanks. Thanks to Ryan for organizing it; and my fellow members of the committee, who I would name if I could remember who else was on the committee; to the Science, Technology and Economic Policy Board, and Gail Cohen, its director; to the Committee on Science, Technology, and Law, and Anne-Marie Mazza, its director; most of all to all the speakers who put on such a tremendous amount to put out, such a tremendous amount of knowledge and questions over the last few hours.

And really, this doesn't get mentioned enough -but to the audience because if you weren't here, it just wouldn't have been the same. So join me in thanking everyone including yourselves. And a good holiday season to you all. We're done.