THE BROOKINGS INSTITUTION

SAUL/ZILKHA ROOM

PREPARING FOR THE ETHICAL DILEMMAS OF THE AI ERA

Washington, D.C.
Friday, September 14, 2018

PARTICIPANTS:

CHRIS MESEROLE, Moderator
Fellow, Foreign Policy
The Brookings Institution

CHARINA CHOU
Global Policy Lead for Emerging Technologies
Google

WILLIAM A. GALSTON
Ezra K. Zilkha Chair and Senior Fellow, Governance Studies
The Brookings Institution

HEATHER PATTERSON
Senior Research Scientist
Intel Labs

DARRELL WEST
Vice President and Director, Governance Studies
The Brookings Institution

* * * * *

P R O C E E D I N G S

MR. MESEROLE: Welcome and thanks to you all for being here, both those of you in the room today and all of those who are able to join us via the webcast, as well. Before we begin the panel I want to make two quick housekeeping points.

The first is that we are joined today by two illustrious colleagues from Intel and Google, so I would be remiss if I didn't disclose that Brookings has received funding from each in the past.

The other is that we'll be opening the floor to questions about halfway through the event.

As many of you are aware, AI is rapidly transforming a number of domains, from education to transportation to national security. And in the process it has raised no shortage of deep and important ethical questions. Thankfully, at Brookings we have two esteemed scholars -- Bill Galston, a senior fellow in our Governance Studies program here; and Darrell West, the vice president of our Governance Studies program here at Brookings -- who have each authored wonderful papers on AI and ethics that are available now on our website.

And in addition to that we're also joined today by Heather Patterson, who is a senior research scientist at Intel Labs, and Charina Chou, who is the global policy lead for emerging technologies at Google.

With that, I'll jump right in to the panel and I'd like to start with Darrell. Darrell, what I admired most about your paper is that you don't just describe the challenges that AI introduces, you also offer a set of policy guidelines and solutions. Would you mind speaking a broadly about why the ethics of AI are so challenging and why they require the guidelines that you lay out?

MR. WEST: Okay. Thank you, Chris. So my paper looks at the role of

corporations in AI development because we're obviously seeing tremendous advances in AI over many sectors: finance, transportation, healthcare, education, defense, and sustainability, among other topics. I have a new book out on the future of work and I argue that people are going to be very surprised how rapidly the AI and machine learning and data analytics revolution accelerates in coming years.

The growing role of AI is going to raise a number of different ethical dilemmas, which I discuss in the paper. This includes the use of AI for weapons development and military applications, its use in law enforcement and border enforcement, its role in government surveillance, the issue of racial bias in AI systems, and social credit systems, kind of online rating systems that measure people's trustworthiness.

Now, some of these problems I think are going to require government action. For example, in the discrimination area I think we clearly need to have very strong anti-discrimination laws specifically applied to artificial intelligence. There's going to be a need for privacy protections and outlining legal rights governing the use of AI in different areas.

But in an era of gridlock and polarization I think it's going to be difficult for political leaders to build majorities in favor of significant action at least in the next couple of years. It's actually hard to imagine Congress passing legislation in the next couple years in these areas or bills like that being signed by President Trump.

We also have the challenge of what in the paper I refer to as dual-use technologies in which many emerging technologies can be used either for good or for ill. For example, facial recognition software can help law enforcement find abducted children, but then it also can empower sweeping intrusions into personal privacy.

So it's often hard to build regulatory regimes in these areas. In the past

it took years to figure out how to regulate airplanes, automobiles, and nuclear energy. And, in fact, now decades later we still make adjustments along the way in terms of how we feel and deal with those types of technologies.

So in the paper I focus on the role of corporations in terms of what I think they should be doing and how they should be thinking about ethics in the deployment of these technologies because I think in the short run this is where the real action is going to be taking place. There are many companies, such as Microsoft and Google, that already put out detailed plans. Intel, Apple, and many others are in the process of formulating their own plans. So in the paper I argue that companies should think about six very specific types of actions in order to make sure ethics are taken seriously in AI deployment.

One, they should hire ethicists to work with their corporate decision makers.

Two, they should develop an ethics code that lays out how various issues are going to be handled.

Three, they should have an AI review board that addresses ethical questions on an ongoing basis.

Four, they need to develop AI audit trails that show how various coding decisions have been made. I think this is especially going to be important as particular problems start to come before the courts. It's going to be very difficult to inspect the millions of lines of code that exist in most software programs, but you can have an annotated audit trail that describes the decision those software designers have made as they were writing the code so that others can understand the choices and that the rest of us can think about the possible ramifications of their decisions.

Fifth, I'd like to see companies implement AI training programs for their

staff that operationalizes ethical considerations.

And then finally, there needs to be a means of remediation in those cases where AI solutions end up inflicting harm or damages on consumers.

So I think all those things would help a lot and they're ideas that can be implemented very quickly. Thank you.

MR. MESEROLE: Thank you, Darrell. I think it's very apt that you talk a lot about the ethics of corporations around AI in particular. Fortunately, we're joined here by an ethicist at one of the leading technology corporations, Intel.

Heather, you've also worked on principles around AI and ethics, and I was wondering if you could lay out some of the principles that you believe corporations should follow, as well.

MS. PATTERSON: I would love to and thank you for having me here today. It's lovely to be here.

And I will say that although Intel is currently working on principles, those are not public yet and so I won't talk about our specific principles. But I will add actually three things, Darrell, that I think could be added to your recommendations. And that may sound surprising for someone coming from a technology company, but I think we can do even more.

So I think it's important to think first of all that AI -- can everybody hear me, by the way? Thank you very much for telling me. So I think it's important to realize when we talk about artificial intelligence that AI isn't one thing. It's a constellation of things. It's a constellation of sensors and machine learning techniques and algorithms and data collection and processing processes. And it's not just something that occurs inside of a computer. It's something that occurs in reaction to, in conjunction with real people. And because this involves real people, the problems are hard. We can't

engineer our way out of them. We need real people to be working on them with us.

And so one of the first things that occurred to me when I was reading your paper is I think it's wonderful to call on corporations to bring ethicists into the room, but I would even step back and say we need to get all of the right people in the room. So not just business leaders and strategists and technologists and ethicists, but also the people who have the ability to do deep dives into the social context in which AI systems and products are deployed, so the social scientists, the anthropologists and the sociologists and the cognitive psychologists and so on. Because in order to really understand the effects of a particular technology you need to look at the social norms that adhere within a particular context, like home life or the workplace or a retail space or a public park or a street.

And really it's important to interrogate what -- a buzzword that's popular now "humancentric AI" -- interrogate what that buzzword means and translate that to design principles. And we're going to see challenges around this, and I can talk about it in a bit, with implementations like the Internet of Things.

Second, not only should the right people be in the room, but those people need to have all of the information that they need in order to make the right decisions. And so I think it's important, companies have an obligation to reduce information asymmetries. And so we talk a lot about transparency, explainability, explicability that refers to also a constellation of concepts, but let's come back to that.

Third, I think it's very important to have accountability and oversight in design and implementation. So I talked about context. It's important that artificial intelligence implementations be contextually and culturally sensitive. Those also need to be backstopped by very clear guardrails, so we know what companies will and won't do. So we can look to things like internationally accepted human rights principles, the fair

information practice principles, rules of war. We're not starting from scratch here, but we do need to be very clear about where those backstops place us in the ecosystem.

MR. MESEROLE: Thank you, Heather. I think where you ended is actually a great point to start with Bill, which is kind of some of the international regulations and norms we've developed. And Bill, if you haven't read it yet I highly recommend Bill's paper. He does a masterful job laying out a lot of the ethics of AI.

And I want to start with one of the points you made in the conclusion, which is that you note that in domain after domain what most ethical concerns over AI share in common is a need for -- or the keen awareness of the limits of self-regulation and the need for government oversight. So what is it about AI in your view that will require that kind of government oversight?

MR. GALSTON: Well, as they say in Washington, I'm glad you asked. I am all in favor of maximum, feasible, ethically aware corporate self-regulation. In my judgment, and it's not my judgment alone as you'll find out in a minute, that is necessary, but it will not be sufficient.

Darrell in his opening statement indicated that it might be a while before legislation catches up to social impact and social change. As a matter of prediction that may be true, but my argument is that the closer we can get to getting out in front of these changes that are going to be so transformative and bring the capacity of democratic decision-making to bear on these incipient social transformations, the better off we're going to be.

Right now, we are still living with the consequences of a profound technologically driven economic transition where we didn't get out in front of the social consequences. And look at what's happened to the politics of democracies throughout the West, in part as a consequence of that.

I said that this view is not mine alone and let me read you a paragraph from a very, very well-crafted memorandum. "Facial recognition technology raises issues that go to the heart of fundamental human rights protections, like privacy and freedom of expression. These issues heighten responsibility for tech companies that create these products. In our view they also call for thoughtful government regulation and for the development of norms around acceptable uses. In a democratic republic there is no substitute for decision-making by our elected representatives regarding the issues that require the balancing of public safety with the essence of our democratic freedoms."

I agree with every word of that paragraph, which was written this July by Brad Smith, who is the president of Microsoft. And for those of you who have not read his memorandum "Facial Recognition Technology: The Need for Public Regulation and Corporation Responsibility," it is I think required reading for people who are interested in this field.

There are a number of different reasons why you need government regulation to backstop corporation self-regulation. One of them, not the only one by any means, is that whenever you have self-regulation there is the first-mover advantage for the person who breaks the norms in one way or another. And that advantage is typically enough to induce one or more actors to deviate from the norms. When that happens, a kind of Gresham's Law sets in where bad behavior drives out good behavior.

A classic reason for government is to have a club in the closet to reduce the incentives for a single agent to break free of binding social norms. We are finding out the hard way in this country what happens when institutions and practices backstopped by norms, but not by law, are violated. And I'm not sure we want to go any farther down that road.

I have lots of things to say in my paper about facial recognition

technology, about autonomous vehicles, and about AI-guided weapons development. Suffice it to say that the issue at stake is nothing less than what kind of society we want to be living in in the next 10 or 20 years. Do we, for example, want to live in a society that is subject to universal surveillance? As we now know, that is not only a technological possibility, but very close to a technological reality.

I can think of lots of advantages. The Brits just succeeded through universal surveillance in identifying the people who are probably responsible for the Skripal affair, although they claim that they were in Salisbury as tourists, which is akin, in my judgment, to claiming that Rick was in Casablanca for the waters. (Laughter) But do we really want to live in that kind of society? Let's call it China for short. I don't, but I think we're going to need affirmative steps by government to prevent that from happening.

MR. MESEROLE: Thank you, Bill. I want to turn now to Charina. And Google has been in the news for a lot of its "AI for Good" efforts. It's also been in the news for things like Project Maven. And in response to that, over the summer it released some principles that it was going to abide by around AI, and I'd appreciate it if you could talk about those.

MS. CHOU: Sure. Thanks, Chris, and hi, everyone. I would say, to set the stage, really at Google we firmly believe that AI is a technology with enormous opportunity for societal benefits.

To give you some context I joined Google almost four years ago and at that time fewer than 5 or 10 percent of our engineers had machine learning training. Today more than 21,000 engineers have been trained with our machine learning crash course at the company. And that course is now available online in multiple languages for anybody to take. I would say this is just one indication of how important AI is to the

company.

And, you know, you think about Google products, right, the opportunity to better understand search queries, the opportunity to serve users in not just English, but in hundreds of languages, to help understand content beyond just text, right? Today a lot of the content there is video and image and audio. It's a lot more complex. So we found that we absolutely need machine learning to adapt to a lot of these questions, better translations and in "Translate." All of that is powered by advances in machine learning.

You know, Google works on a lot of hard problems, but it's really only a fraction of the problems that exist in the world today. So if you think about translations that can maybe help doctors communicate better with patients who speak another language or image recognition for earlier detection of cancer, right, these are all things that we see as enormous opportunities.

Often if we get the question is it ethical to pursue AI, I think the flip side of that question is given all of these potential opportunities is it ethical not to pursue AI? Right? What are all the potential benefits we would be foregoing if we did not pursue it? So for Google, the clear answer is that absolutely should, need to, want to pursue AI.

That being said, I think it's not enough to just lead in the development of AI technology. As Chris noted, and others, there are many potential risks and consequences if the technology is developed in the wrong way or possibly applied to harmful purposes. Something that we are doing, our approach to self-regulation at this point is really thinking through what does it mean not just to be a leader in AI technology, but also a leader in the responsible development and use of the technology? And that was really the context for us to put together a set of Google AI principles. You can read them on our website.

There's two reasons for this, two purposes for this set of principles.

Number one, of course, publicly it's our way to share externally how we're thinking about, how we're approaching AI technology. But also, number two, internally it's our way to assess various features and products and deals and applications and new ideas, even research that we might be pursuing. Every new opportunity we are using these AI principles as a framework to assess various options and see whether or not this is in line with our principles or not.

In addition to our principles in which we lay out seven things AI should do or be and four red lines, we also released a set of AI responsible practices. You can also find this at the Google AI website. This is a very detailed set of technical practices. A lot of times people say great, you know, I also care about fairness and privacy and accountability and safety. These are all really good things, but how do you actually live up to that. Right? To Heather's point, how do you build that into your products? What are technical steps that you can take? And we have very detailed technical steps at our responsible AI practices.

The idea there is Google's doing a lot in AI, but actually right now a lot of people are using AI. Right? The barrier is getting lowered. There are so many opportunities for people to use AI for the problems that they care about. They should also have responsibility for the things that they're developing. And what we can do at Google is really share the best of what we know.

You'll notice that they're called "responsible AI practices," not "best practices" because we don't think that we really know best yet at this point. To be honest, we don't think -- it's early days, we don't really think anybody knows best. But we are keeping that updated every quarter so we really can share the latest of what we know and what we recommend at this point.

Of course, we're thinking through internal governance. We have a lot of

things that we're working on there, how we actually implement the principles at Google. And maybe the points I'll leave this on is how we really just are one of many players, right? I think there are a lot of voices that need to take part in this conversation. But as Google, we absolutely want to collaborate on doing our part to develop this in a responsible way.

MR. MESEROLE: Thank you, Charina. It's been a wonderful round of questions and remarks. I'm going to exercise a bit of moderator's prerogative and ask a couple of follow-ups before we turn it to the audience.

Darrell, I wanted to start with you, which is that I think one of the most interesting things about reading your paper against -- and reading Bill's at the same time is that there's this tension between, on the one hand, we want to rein in the abuse of AI by companies and so we want government to have more oversight; on the other hand, we're also very much concerned about the abuses of coupling -- or the possibilities that emerge when you couple AI with state power. And so I'd be curious for your thoughts on how to navigate that tension.

MR. WEST: Okay. Great question. Thank you very much.

I actually agree with Bill that AI does pose a particular risk. I mean, he gave the example of facial recognition software and how some countries around the world are really using this in detrimental ways. That clearly is a very problematic subject, especially in the hands of law enforcement, so there needs to be some government action there.

Ditto in terms of racial bias types of questions. We know that at least some, perhaps a lot, of the data being used to train AI systems either are incomplete or outright biased. In facial recognition, for example, AI is much better at recognizing Caucasian than minority faces. So, for example, for darker-skinned women AI image

software has about a 35 percent error rate; for darker-skinned men it has a 12 percent error rate. The Caucasian rate is much lower than either one of those figures. And so that kind of illustrates how data biases can become AI biases. And so we really need nondiscrimination principles applied to various types of emerging technologies.

But when you incorporate the governance angle in these types of questions and given the very difficult political situation that we face now, especially at the national level, you know, I think it's going to take several years to even get to the point where it may be possible to pass legislation through Congress, get a President to sign this. I don't think we have that much time.

When you look at how AI is already being deployed in terms of autonomous vehicles, like self-driving cars are not about the cars. It's about the AI. The AI is what integrates the LIDAR information, the imaging data, the sensor information in real time so you don't have an accident. That stuff is going to be on the road starting next year.

Finance software is deploying AI. Healthcare is deploying AI. So I think companies have to be much more transparent about how they're thinking about AI development and deployment. They have to be transparent about how they're making the decisions and how they are incorporating ethical values in their decisions.

We've actually done a series of monthly public opinion surveys on public attitudes towards various emerging technologies. So, for example, in May, we published a survey about attitudes towards AI and we found considerable fears about the AI impact on jobs, on personal privacy, a worry that robots and AI are going to be taking over from humans. Companies have to take these fears seriously so that they do no provoke a tech lash, kind of the backlash against technology that we already see starting to emerge in society as a whole.

If the public concern becomes too widespread it's actually going to guarantee very tough government regulation down the road. So companies better take this seriously if they want to maintain the public trust.

MR. MESEROLE: Thanks, Darrell. I want to pull off on a couple of the points you made about bias in particular and ask Heather, you know, one of the things that you work on is explainable AI and how that form of AI can potentially mitigate some of the bias issues.

MS. PATTERSON: Thank you. I think it's an incredibly important point that we need greater transparency and greater explainability where possible. And so to that end I would say, you know, there are some types of AI where explainability at this point from a technical standpoint just isn't possible. But that doesn't mean that we can't invest significant resources as a society, but technology companies in particular, in finding ways to make more -- to make clearer what parameters are being assessed by a particular machine learning model and how those parameters are being weighted to make decisions, so that we don't have problems, like facial recognition algorithms not recognizing brown-skinned people as accurately as white-skinned people. That's hugely problematic.

We need to understand other things, like how do we let individual users of technology know, understand when it is they're being monitored, when their information is being collected, how it's being transmitted to, and for what purposes? We haven't built a lot of that into technology right now. But looking forward, we may have an opportunity to do so.

So explainability, you look at explainability from the aims and purposes of the system to the quality of the data that's being collected, making sure that the representation in the database that the machine learning models are using is appropriate

for that use. And thinking also about the effects that our technology, all of our technologies, are having on real-world communities post deployment.

So we put all of that into the bucket of explainable AI and say we do need to be having this conversation. We do need to be working with parties across sectors. And we do need to prioritize this as a society.

MS. CHOU: Could I add to that really quick, Chris? I completely agree with Heather. I think fairness and explainability is crucial for AI systems.

I did want to make the point that actually many of these machine learning systems can help us to be, you know, even more responsible than we are today as humans. Right? It's important to develop them responsibly, but also they can help us to be even better.

So in the case of fairness, a few years ago Google worked with the Geena Davis Institute to look through thousands of hours of movie and television footage and they had one question: How often do men have screen time and speaking roles and how often do women have screen time and speaking roles? We partnered with them using a Google machine learning model that's able to identify male faces and female faces to automatically annotate the thousands and thousands of hours of footage. Right? That's something that would not have been possible without a machine learning model. Nobody's going to sit there and just mark it.

But what they found is, maybe not surprisingly, men have a significant majority of speaking time and screen time relative to women. And what was a little bit surprising was actually that in movies that had female at least 50 percent of the screen on speaking time, if not more, they actually performed better at the box office.

So these are real stats that were made possible through the use of machine learning. I think that's really important to note.

On the idea of explainability I think the same thing. Right? It's absolutely important to have explainability in our machine learning models. But actually they might be able to help us explain more about decisions than we can today as humans.

My husband is an oncologist, so he sees cancer patients all the time. And sometimes, you know, it's very clear what they have, but sometimes he just has a feeling. Right? He can say I sense that maybe you should go back for another PET scan or you should maybe get rechecked. I feel like the cancer is metastasizing, not really sure, not really able to pinpoint it exactly. Usually he's right, but it's really that intuition with machine learning systems. And it's still early days, but they're getting better.

There's actually the opportunity to identify why a certain decision is being made. Right. Today especially there's been a lot of progress in explainability of machine vision systems, being able to look at an image and say, all right, I think there's a detection of a tumor, but not just yes or no. Here's the section where I think tumor cells are existing and here are the previous 10 examples that this model saw before which is leading it to this decision.

So I did want to make that point.

MR. MESEROLE: Thanks so much for jumping in.

MS. PATTERSON: I just want to build on that a little bit. I absolutely agree. I think we also need to make sure that professionals like oncologists and others do have the opportunity to weigh in, so that these decisions -- when we think about deployment of AI technologies the decisions aren't just made by themselves and they don't just necessarily take effect, at least until some generally accepted level of accuracy has been achieved and we're all comfortable with that as a society. We don't want to cut out human expertise when there may be -- we may need to build in opportunities for the

humans who are professionals in their own domains to take a look at those scans and say, hey, actually, you know, this the system is indicating that there's a tumor here, but I think it's something else and I think there might be something going on over here. So remembering that these technologies are being put into place to help people, to serve people, and not to replace people.

The second point is let's just keep in mind, as well, that implementation is important. So when you think about algorithmic bias and, say, facial recognition technologies perhaps not being accurate enough or, you know, the flip side of that great accuracy might, in certain implementations, actually be problematic. So we talked about surveillance states and autocratic regimes. There might be situations where you want to be cautious about how particular technologies are actually being deployed and make sure you follow up and remain aligned with the principles that you've set forth as a corporation or as another actor.

MS. CHOU: Completely agree.

MR. MESEROLE: Problem solved. (Laughter) Before we turn it over I want to ask one last question to Bill, which is that we've talked a lot about the ethics of AI. We haven't quite talked that much about moral agency and moral frameworks that we use to answer those decisions.

One of the things that I liked about your paper was, you know, looking at things like driverless cars and autonomous weapons, you put the moral agency question to some extent front and center. And I'd appreciate your thoughts on, you know, especially who has moral agency with these systems and how it should be used.

MR. GALSTON: Well, that is a deep question that I can't possibly answer in a minute or two or three or five or the remaining time for that matter. So suffice it to say that I tried to put the problem on the table by working through a series of

examples.

With regard to driverless cars, here's my example. The car is driving itself along and then suddenly there's a bouncing ball that comes into the road followed by a four-year-old running after the ball. Both the human and the autonomous vehicle guidance system would reach the judgment in this case let's say I can't break in time to avoid hitting the boy, so what do I do? Right? And I worked through a series of examples.

And one of the points that I make is that what you do may be dependent on who you are, who the other passengers in the car are, you know. For example, if you have your two children in car seats in the back and you know that if you swerve left and there's an oncoming car, there's a pretty good chance that your car is going to be broadsided and risking harm or even death to your own children. All right. What do you do?

In the case of AI, what does who do? Right? How is that autonomous system programmed to deal with morally fraught situations of that sort? And I don't want to sit here and give a glib answer. But I will say that the displacement of that decision from the mother driving the car to some designer, anonymous designer someplace, is not a transfer that I'm comfortable with without having substantial assurance that the sorts of ethical considerations that a parent would bring to that decision have been fully reflected and to a higher degree of accuracy and efficacy than the human mother would be able to reach. And I am far from persuaded that either of those conditions has been satisfied or will soon be satisfied.

MR. MESEROLE: With that, I will turn it over to the audience for Q&A. If anybody has any question, please identify yourself and state your question as concisely as possible.

Yes, sir?

MR. BIGGS: Zach Biggs. I'm a national security reporter with the Center for Public Integrity.

So I wanted to ask, and sort of circling back to the prior mention of Project Maven, but broadly working with the Pentagon as tech companies and working on the AI problem with tech companies. We've seen a little bit of a history of companies being hesitant, whether it's what happened Boston Dynamics, whether it's Project Maven, even with technologies that aren't inherently part of a weapons system.

So I guess my question, particularly for the two members of the panel from industry, is the problem who the customer is when it comes to working with AI and the government? Is the problem that it's the Pentagon, regardless of where the AI's being used, or is the problem the actual application of the AI technology?

MS. CHOU: I'll step in first given the question and specific pieces of it. I will say on Maven obviously we have everything that we have to say publicly written in blog posts by our CEO Sundar Pichai and CEO of Google Cloud Diane Greene. But beyond that we outlined our AI principles to really serve as a guideline, a gold standard, for how we will consider the various projects we work on.

You'll notice we had seven principles that outlined things that we believe AI should do or should be. We've also outlined four red lines. One of those red lines is that we as a company will not work on AI with applications to weapons. So in the case of the specific question that you raised it really is in the application of it.

We've emphasized multiple times, and this is stated clearing in Diane Greene's blog, as well, that we are absolutely committed to working with the military and the U.S. Government. I think we have a lot of ongoing collaborations already. There's a lot of opportunities for continued work in things like search-and-rescue, in healthcare, in

education, and disaster relief. So we're absolutely committed to that.

MS. PATTERSON: And I'll also start by saying I'm a social scientist at Intel. I'm not formally on the policy team, so I can't speak to that. But I will say a couple of things.

I do think that we're having a moment right now in society. And so I think it's interesting that we're seeing among employees of technology companies a lot of interest internally about how their companies are behaving, who they're engaging with, how technology is being deployed, and the effect that it's having on real people in real-world communities, and a real sensitivity to that.

I don't know that we would out of hand not engage with particular customers because of a position they hold in society. My hunch is that we would look to the principles that we have internally that we're developing, but also our longstanding position as a very ethical company, and reflect on those and choose projects that are in accordance with those guidelines that we have always followed.

MR. WEST: And if I can add a public opinion component to your question, we recently did a survey looking at public attitudes towards AI for warfare. So it kind of gets at the general thrust of your question. We found only 30 percent of Americans right now support the development, the deployment of AI for purposes of warfare. But then we asked a follow-up of if we thought our national adversaries were developing AI for warfare, you know, how would you feel about it? The 30 percent support jumps up to 45 percent support. So how people feel about these issues is not just dependent on either the customer or the application, but how they see the general situation.

We also found very significant age and gender differences in how people responded to that, with men being much more supportive of AI for warfare than women;

and older people, especially senior citizens, compared to younger people being supportive. So I thought those were kind interesting angles on that.

MR. MESEROLE: Great, thank you. Yes?

MR. PEHA: Hi, I'm Jon Peha, Carnegie Mellon University.

It's great that companies have general principles, but consumers generally don't know what actual decisions are made on AI issues, nor do shareholders, nor do most employees. Is there any example of a company that has done a good job of being transparent about decisions actually made without a lawsuit or scandal?

MR. WEST: No. (Laughter)

MS. CHOU: Well, I'm going to jump in and disagree with that just a bit. Could you clarify a bit what you mean on types of decisions? Do you mean decisions in products or decisions about technology in general?

MR. PEHA: I mean, well, technology products, products or services, where you've had to make value decisions applying some of the principles that we are talking about. And you have let relevant stakeholders understand the -- actually not just you, anybody, it doesn't have to be your company, but is there examples where meaningful transparency is there so people understand the hard decisions, not just the easy ones or the general principles, that have been made?

MS. CHOU: So I'll give a few examples, one on our Search product, for example. Anybody can actually go on Search. There are Search rater guidelines, clear explanations of how Search quality is defined. And that dictates the results and the processes by which things come up in the Search platform. I think there's a lot of opportunities like that in other products by other companies, but I'd say in terms of decisions that we've been making with the AI principles that we set out, we absolutely wanted to make those principles public so that everybody could see the types of

principles and high-level guidance that we're living by.

As I mentioned, we also have the responsible AI practices, which you can check out. They have many, many, many examples of very technical details and decisions that we have in terms of fairness and explainability in privacy and in security. So those are all there.

The last thing I would say is, you know, we're thinking through a lot of new issues that are coming up, right, with the advent of machine learning and all these new opportunities; obviously, also comes new questions. And they are things that I don't think many people have asked before or have had to ask before. So as these come up, we released our principles a little more than three months ago, I'd say in the three months since then we've had many different kind of submissions and things to look at in terms of potential research papers and features and products that we might want to launch.

We're definitely taking a very close assessment of each of those with respect to the principles and we'll be figuring out kind of the best way to also share and disseminate that information in a way that's made public.

MR. WEST: And if I can just quickly follow up on that. I mean, when I look at the ethics codes that are coming out from a variety of organizations, not just businesses, but NGOs, universities, and so on, like I don't disagree with any of the principles: fairness, safety, transparency. We're all in agreement with that. But I think your question really pinpoints the nub of the problem, which is how do those very general and abstract principles get applied?

And I'm sure companies are having a lot of discussions about this. They're actually making decisions. But as you suggest, we don't know the results of those decisions. So I think companies are going to have to do better if they want to

reassure a general public that is becoming skeptical of a range of new technologies, not just AI, but driverless cars, robots, facial recognition, and a bunch of other things.

MS. PATTERSON: I agree. And I think you raise a really important point. It may be that companies have not done a fantastic job about communicating their internal policies and practices up until now. But I think we are seeing a really concerted effort to be better at this. And so when I talked about transparency and all of the different layers of that, one of those is communicating internally and training employees, socializing them to the ethical principles that are being developed that we want to see practiced; communicating to shareholders, communicating to policymakers, but also communicating with civil society organizations and making sure that they have reviewed principles and practices and making sure that there's buy-in, but also communicating with each other.

So there are bodies, like AI Now and the Partnership for AI, IEEE, and those are all fora for technologists and business leaders and ethicists to get together to really talk through these problems, not just the development of the principles, but the operationalization of those principles in practice. And so one place to look is, for example, IEEE's ethically aligned design whitepaper that they've published two versions of now. And that is a 250-something page paper that lays out most of the ethical issues associated with artificial intelligence now.

There's been a call, the public comment call may be over, that window, but a call for contributions from people throughout the world. And there's an interest in translating those into a series of standards, the P-700 -- the P-7000 series of standards for how AI gets deployed.

So those resources are available. It's our goal to make them more accessible to everyone.

MS. CHOU: Yeah, and as it's been noted, a lot of these principles are very high-level. Right? Who is going to disagree with privacy or security or fairness at a high level? So a lot of the details are what are going to shed light on kind of the decisions being made. I would say, of course, technical details on implementation, but also what happens when some of these principles come into conflict with each other. Right? How do you weigh them against each other?

For us, we really wanted to write a set of principles that was very high-level in terms of allowing room -- we don't know exactly what the technology's going to look like in 1 year or 10 years. Don't want to be too prescriptive in terms of what's dictated. But, at the same time, we want to give good guidance. So something that we've been looking at, as I mentioned, in the last three months, and we'll see how other companies and organizations do this, too, is to really develop a case law of sorts, examples, to say like here's how we've interpreted it in this way and here's how we interpreted it in this other situation to start building an example and a record of decision-making.

MR. GALSTON: If I could just jump in here for a minute. The fundamental question underlying the considerations that have just been put on the table is who decides? It is absolutely the case that high-level general norms are going to come into conflict almost inevitably with something that is going to have a range of effects. Who decides which one of those norms takes priority or how they're to be balanced?

If you draw up, to vary the decision, the metaphor just a little bit, suppose you draw up a list of what you consider to be social benefits and social costs from the development and deployment of a new technology.

Let us say further what is almost always the case, these costs and benefits are -- you know, they don't have a common denominator. They are qualitatively

different. Who looking at those two lists decides which way the balance implies?

There are some circumstances in which individuals make that decision, some circumstances in which the private sector makes those decisions. But for the kinds of issues that we're talking about, I think both history and logic suggest that there's also a role for the public sector making these decisions. So we talk about explainable AI. Explainable to whom? Right?

And I don't think that question, when posed, answers itself. Explaining to consumers is not necessarily enough because of tragedy of the commons problems, that is a transaction that is both explainable and attractive to an individual consumer, when you multiply that by 100,000 or a million generates a new social reality. So, for example, New York City is now grappling with the fact that the proliferation of Uber and Lyft is contributing to traffic congestion, further traffic congestion I shall say, which is measurably slowing traffic in New York City. That is an aggregate social consequence of individual, technologically driven capacities that didn't exist 10 years ago. And I don't think that individuals or Uber or Lyft are going to decide the question as to whether that's an acceptable social outcome or not. At some point New York City's going to have to make that decision, public servants, elected representatives with all of their flaws.

That's the only point I'm making for this entire discussion, you know, pretending that self-regulation is going to substitute fully for public judgment of this sort is I think a fool's errand. And that's the only point I'm making.

MS. CHOU: Yeah, I would -- oh, sorry. Yeah, I would just say completely agree with that. I think there are a lot of different voices that need to be part of this discussion, including public sector. Right? There's a lot of these shared decisions that we need to make together.

I will say kind of as different players are moving in organizations, some of

them are taking time, right -- what we have as a company at Google is opportunity to move on our own and we are doing what we can in this area.

MS. PATTERSON: Right. I think we also view this as doing our part in the ecosystem and not at all intending to exclude any other actors and, in fact, inviting other actors to participate in the conversation with us, please.

Just for one example, you mentioned New York City. I know that they're actively working right now with their -- they have an algorithmic decision-making task force that is going to decide -- I think that they're tasked with understanding how potential AI implementations and adoptions within the city of New York, how that is working in practice. Who will be affected by those technologies?

And so we think about explainability, explainability not only to technologists and to individuals within companies and end users, but explainable to governments, not just regulators, but cities and counties and states who may actually want to be purchasing and using these sorts of technologies. They have an obligation to the public to be able to understand, for example, how bus schedules get set or how public resources may be distributed or how sentencing guidelines might be decided upon.

So absolutely in agreement that we all need to be talking here.

MR. MESEROLE: Oh, boy, the hands have come up. I'm going to take three quick questions and then try to keep your questions very quick. And then that'll be it and we'll have to wrap up. First, right here.

MS. OTURNO: Thank you. Yamin Oturno with the Interactivity Foundation.

We organize public policy discussions among the general public and AI is one of the areas we just started looking at. So my question to any of the panelists would be what questions you would suggest the general public should be discussing

when they're looking at AI, not so much today, but 5, 10, 20 years later? Thank you.

MR. JANG: Hi there. Jae Jang with the American Bankers Association.

So I wanted to ask about the fair lending and UDAP risks associated with the employment of artificial intelligence. You know, so for example, in financial services an area that we're very concerned is where AI plays into underwriting or marketing and pricing. The fear is in a way where AI can be used to limit offerings in a discriminatory way for people.

So my question, whether it's in banking or more broadly, the deployment of AI across industries, is how can we ensure we have transparency and fairness? And on the other side, you know, how can we prevent AI from really exacerbating and making worse let's say the implicit biases that we tend to have individually? How can AI -- how can we think about the issue? Thank you.

MR. MESEROLE: I'm actually going to wrap it up with those two and give each of the panelists an opportunity, hopefully in about a minute, to answer one of the two questions. (Laughter)

MR. WEST: Well, I can jump on the finance question that this gentleman just asked. I think you're right in that AI does have the potential to exacerbate unfairness, inequality, and a host of -- as well as bias and discrimination issues. The great fear is that we will end up with digital redlining, that all the digital data actually will make it even clearer in terms of problems that they're doing.

That's the reason why I think in these types of areas we do need nondiscrimination legislation to basically say you can't do it that way. And if companies are found to have violated that, in the same way that banks in the past engaged in traditional redlining, then they are subject to civil or even criminal action.

MS. CHOU: Sure, I'll answer the first one on general public. I would say

obviously there's some initial questions around how do you, you know, speaking to your audience feel about things like fairness? Where do you draw the line, right? Even when we talking about fairness, there aren't clear definitions for exactly what that looks like. And I ask myself versus you versus other people in the audience, we might all have different explanations of what is fair for a given situation, right? So how do they think about that? How should those types of decisions be made available, be make accountable?

The same thing around explainability. What level of explainability is appropriate and for what types of applications? For example, if you use a product like Google Translate, sometimes it comes up with -- most of the time it's pretty good, but sometimes it comes up with funny translations. And in that case, you know, usually you just kind of laugh and it's funny. But it's not as funny if it were in a case like a financial credit decision or a medical decision, right? In that case you probably want a much higher level of explainability.

So I think it'd be really valuable for the public to be thinking about these things and impacts. Where do they draw the line? How would they assess this? What would be most useful?

And then I think the last thing I'd note is it'd be amazing if the general public was asking the question around how can I get involved? Right? We talk a lot about AI for everyone, this opportunity that AI can bring. The last thing that we would want, of course, is for this to be something that increase inequality, right? There's so many opportunities, the barriers getting lower.

Right now I don't know how to build a car. Probably most of us in the room don't know how to build one, but we can certainly drive one. We can operate one to get from where we need to go, point A to point B. So same thing with AI, there's

already kind of starting to be great models and pre-built APIs and things like that, but even better interfaces. How can the public get involved? What are the problems they care about? Even with government using AI to develop and deliver public services better, I think there's a lot of opportunities there.

MR. GALSTON: Very quickly, your banking and finance question raises a really profound issue, which I think ultimately will be a philosophical issue. And that is, is human judgment subject to algorithmic representation and reduction? I think the jury's very much out on that.

And one of the things that we already know is that the effort to reduce lending practices to quantitative metrics has had the effect of disadvantaging borrowers, particularly small borrowers, who used to be the beneficiaries of what was known as relationship lending, where people received loans based on a loan officer's personal knowledge of the character and history of the prospective borrower even if that borrower couldn't meet standard income and collateral standards.

And I think the step towards standard quantitative metrics for lending is clearly having asymmetric effects for smaller borrowers and more remote borrowers who are increasingly distant from increasingly concentrated financial entities as community banking disappears, for example. No relationship, no loan.

So the question of human judgment I think cannot be taken out of the equation.

MS. PATTERSON: I will also answer the first question. So by way of analogy I think about privacy. It's been challenging to figure out how individuals think about privacy, how much they care about privacy, and how that gets manifested. So a popular point of rhetoric is that individuals don't care about privacy, privacy is dead. Privacy researchers know that that isn't true. We know that when you look in, as I've

said, particular contexts -- the home, transportation, the workplace -- people care very deeply about what information is being collected about them and how it's being used, who it's being transmitted to, and what control they have over that kind of information getting out.

So similarly, it's difficult I think -- I also want to know more about how the public feels and I think it's really important for all of us to be asking and answering questions. It's difficult to say how do you feel about artificial intelligence? I think it's important to be very specific.

And thinking about, you know, even something like face recognition technology, even within a particular context, somebody might feel that it's fine for face recognition to be used in lieu of badging into their workplace. They might be wowed by the ability to walk up to an elevator and have that elevator take them up to the sixth floor where they work. They might be less excited if they realize that that same technology could be used to track their presence across the building, potentially evaluate their performance, their efficiency, and maybe with dire consequences. Maybe that could be linked to their pay or their chances of promotion.

It's important to develop case studies and be very careful about the analytical framework of those case studies to make sure that you're really pulling out what it is making people feel uneasy. That's information that I think that we all need to have as we develop these technologies going forward.

And I would also say it's important for all of us to realize that artificial intelligence is personal. The data that is being collected, that is in databases, that is being used to train machine learning algorithm, is about us. We're the ones who are supplying the data. We're the ones who are using the personal assistance and the smart lights and the smart lightbulbs. And without our information this is not going to happen,

it's not going to work.

So we need to all be aware of the role that we are playing in the ecosystem and decide whether we want to play that role and to what extent we want to do so.

MR. MESEROLE: Thank you, Heather. That's a great note to end on. I want to thank you all for being here and for offering your wisdom and expertise.

I want to thank you all, as well, for joining. And we didn't quite solve the problem necessarily, but I think we got a good head start on transparency and explainability and accountability, so thank you all. (Applause)

* * * * *

CERTIFICATE OF NOTARY PUBLIC

I, Carleton J. Anderson, III do hereby certify that the forgoing electronic file when originally transmitted was reduced to text at my direction; that said transcript is a true record of the proceedings therein referenced; that I am neither counsel for, related to, nor employed by any of the parties to the action in which these proceedings were taken; and, furthermore, that I am neither a relative or employee of any attorney or counsel employed by the parties hereto, nor financially or otherwise interested in the outcome of this action.

Carleton J. Anderson, III

(Signature and Seal on File)

Notary Public in and for the Commonwealth of Virginia

Commission No. 351998

Expires: November 30, 2020