THE BROOKINGS INSTITUTION

FALK AUDITORIUM

# WAYS TO MITIGATE ARTIFICIAL INTELLIGENCE PROBLEMS

Washington, D.C.

Thursday, October 31, 2019

PARTICIPANTS:

DARRELL M. WEST, Moderator
Founding Director, Center for Technology Innovation
Vice President and Director, Governance Studies
The Brookings Institution

ROBERT D. ATKINSON
President
Information Technology and Innovation Foundation

NICOL TURNER LEE
Fellow, Center for Technology Innovation
The Brookings Institution

JOHN VILLASENOR
Nonresident Senior Fellow, Center for Technology Innovation
The Brookings Institution

* * * * *

P R O C E E D I N G S

MR. WEST: Good morning. I'm Darrell West, Vice President of Governance Studies and Director of the Center for Technology Innovation at the Brookings Institution, and we would like to welcome you to our event on Artificial Intelligence.

So AI is advancing rapidly. It's powering new applications in health care, finance, and transportation as well as a number of other areas. Cities are starting to deploy AI design to help with parking, traffic congestion, and air quality, but at the same time there are concerns that AI is undermining fairness, creating privacy difficulties, and threatening human safety. People worry about a loss of human agency over this and other kinds of digital technologies.

To help us understand these issues, Brookings is launching a new AI paper series today. We published three papers this morning and there will be more in coming weeks. There's a paper by John Villasenor; you'll hear more about that today on products liability law as a way to address AI harms.

My colleague Tom Wheeler has a paper out on the lessons of history for regulation of AI, and then I have a paper entitled Ten Actions That Will Protect People From Facial Recognition Software, and I won't go through all the details on my paper, but I suggest things such as time limits on data storage providing clear notification to the public in areas where facial recognition is being deployed creating accuracy standards for facial recognition, deploying third-party assessments, developing technical standards in making sure that facial recognition is based on representative data.

So you can find it as well as the other papers online at Brookings.edu. Overall we will have around

two-dozen papers by prominent experts both inside and outside of Brookings. These will be in coming weeks, so we tried to find the very best people who could help us explain AI to a general audience and make recommendations for mitigating possible problems.

So today, we have brought together three distinguished experts. Nicol Turner Lee is a Fellow in the Governance Studies at Brookings. She writes about digital disparities and new developments in technology innovation, and she is organizing a series of papers on AI biases and these papers will be coming out over the next few weeks.

John Villasenor is a professor of engineering and public policy at UCLA and he's also a

nonresident senior Fellow at Brookings. He's an expert on digital technology, deepfake videos, and AI, and the author of the products liability paper that we'll be discussing in a minute, and Rob Atkinson is president of the Information Technology and Innovation Foundation. He covers a wide range of technical issues including AI, and he has a forthcoming paper in our series on ways to distinguish reasonable from unreasonable AI biases, so we will get into that as well.

But I think I will start with Nicol. You've written about problems of bias in AI and ways to deal with that. What kind of AI biases should we be worried about?

MS. LEE: Thank you. Darrell. Thank you, everybody, for coming out today. We're really happy to have you here and I want to echo Darrell that we're really excited about this paper series, so when my particular paper series releases on bias, you're going to see a series of papers written by academics, other experts both in the engineering field as well as in the social science field, that will speak to levels of bias as well as unravel, I think, as a sociologist things of importance to me which are measures of disparate impact or some of the structural societal consequences that come out of bias either at the beginning or the end.

And so where are we with bias? And so I see one of my coauthors of a paper that we put out at Brookings early this year, Genie Barton. We actually really think bias is one of those strange places and strange place -- areas to define. Because there's always some level of discrimination actually occurs on the Internet, whether or not I like tan shoes over red shoes or this beautiful blue dress that I'm wearing today, the Internet knew that this was my preference over some other color, that type of discrimination in terms of our preferences in lifestyles happens most of the time.

And I say to people, that 80 percent of the time where those innocuous searches, search queries deliver those kinds of results to me, they're probably harmful. I probably have more blue dresses in my closet than I want, but guess what, that's what I like.

The challenge in that 20 percent is when you start seeing AI applied in areas like criminal justice, employment, financial services, health care where we actually see greater problems. So some of you who read the newspaper read about a recent algorithm in health care that biased against African Americans. Because of the risk score that was assigned to them, those patients who had a tendency to have higher levels of chronic diseases, whether high levels of diabetes or blood pressure were sort of

assessed out of the model. So 50 percent of those patients were not qualified for a special program to sort of help discern and unpack those diseases and give better care. Why is that a problem?

It means that that group of people who happened to be African Americans who were part of this biased health care algorithm remained sicker and do not have improved quality of life. In criminal justice, we see similar types of consequences. We know that algorithms are deployed, particularly the COMPAS algorithm, to create levels of efficiency to streamline the burdensome case work that judges have when it comes to bail and sentencing. But when an algorithm is based on training data that represents the United States of America's fractured criminal justice system where African Americans are disproportionately represented as criminals, then you over-criminalize people in the algorithm because it pulls the data which is most represented on both sides.

And so what I suggest to all of you when we talk about bias, I can keep going because Darrell knows I'm like a preacher on a Sunday when it comes to bias. I can talk about in education; I can talk about it in employment. The challenge that we're seeing in algorithms is this conversation around the objectivity of the math, but the disconnect between the context in which it's deployed, and that bias, even though it may appear to be, you know, causing minimal harm, has the consequence to cause greater harm thereby amplifying the stereotypes of discrimination that we actually have today.

And that's, again, one of the really great areas of the research paper series that we have, not to do a shameless plug, but we're really interested in unpacking then, what do policymakers do when they have systems in place particularly AI systems that have not necessarily been litigated in court or actually pushing up against areas like the Fair Credit Act, the Employment Act, that have already been pre-settled prior to the advent of the Internet. Who do people go to? And so again, this untangling and unpacking of what bias looks like I think is a really interesting area and we're only at the beginning.

MR. WEST: Okay. Thank you. So, John, Nicol just mentioned the role of the courts, so which fits nicely into your paper on products liability Law because it turns out there are legal cases that are starting to percolate and judges are going to play a very important role in litigating these questions and helping to resolve them, and your paper looks at consumer harms, how the law can help mitigate AI problems. Can you tell us how products liability law can be helpful in thinking about AI?

MR. VILLASENOR: Right. So as all of you I'm sure know, products liability is the area of

law that deals with harms, either harms to persons or to property that are caused by defective products, and the intersection of products liability law and AI is one of the many extremely important and fascinating intersections. There isn't a lot of case law because there hasn't been a lot of cases, but there has been an enormous number of cases in product liability, generally, over the decades and so it's a very well-developed field of law.

There's actually an interesting mix of torrent law and contract law. Torrent law -- you're all familiar with torrent like in negligence and things like that or design defects, or manufacturing defects, and then contract law is implicated because in any sale, there are implicit -- or explicit warranties that are made by the buyer to the seller, so one of the really interesting questions is, if you take this very well-developed body of law and then you apply it to the inevitable cases where an AI algorithm is going to cause harm, whether it's because it wasn't designed well enough, whether it's because the data caused it to do something that the original designer didn't do, the question is what to do about that.

So on the positive side, I'm optimistic that while there isn't a lot of precedent, there's a lot of very good -- there's not a lot of precedent specific to AI in products liability, but there's a lot of precedent in products liability in general that can be brought to bear, I think, not perfectly, but generally effectively in the AI context, and products liability is one of these areas of law that has been very dynamic and adaptive over the decades in response to new technologies, and I don't see any reason why it can't also be adaptive in respect to AI.

One other overarching conclusion that I believe, and I express in this paper, is that companies need to bear responsibility for the algorithms that they create. So in other words, if a company, you know, ships an AI algorithm, and AI, as all of you know, one of its distinguishing attributes is that it can evolve.

Like the algorithm itself, you know, can create new versions of itself that maybe no human programmer ever sat down and typed out, and you might imagine that companies might offer defenses like, well, you know, when we launched the algorithm into the market it was fine, but it evolved in ways that, you know, that had nothing to do with us. Well, no, I don't think that's a legitimate excuse. I think if they made the algorithm originally and they sort of gave it the framework within which it would evolve, then they have to own the problems that arise later on.

You can also imagine that companies might blame the data, right? I mean, the last thing a company that's going to sue is going to do is say, yes, you're right and how much money do you want, right. You know, they might blame the post-sale evolution, they might blame the data; say, well, the algorithm's great, but, you know, the data, they might blame the user, right, and there will be in some cases, you know, be examples where a user of an AI algorithm is actually, you know, somebody took an algorithm that was intended to diagnose MRI images, and they applied it to CT images, and they didn't get the diagnosis -- they missed the diagnosis, so at that point, I think there's not really a products liability claim. But in many cases, there will be some responsibility, and I think it'll be important to attribute it back to the company. The final thing I'll say is that sounds like a nice clean sound bite, but the challenge is the supply chain is very complicated, right. It's not always the case that a single company conceives of a product, wrights the algorithm, and then ships it out there and that's it. Very often the algorithm is an amalgamation of code from multiple sources and the question of attribution which has always been at the center, of course, of products liability claims becomes very significantly more complicated in some cases with respect to AI. So fascinating area, important area, and one that I think can create a set of incentives that will help suppliers of AI Solutions try to avoid solutions that could cause harms.

MR. WEST: Okay, Rob, you have a forthcoming paper in our AI series along with a couple of ITIF collaborators on how to distinguish reasonable from unreasonable AI biases. It's a very interesting paper. We'll be publishing it in the next couple of weeks. What are the differences there; what are the kinds of biases that should be acceptable to the general public?

MR. ATKINSON: Sure. Well, thank you, Darrell, for having me here. You know, I think one of the things would be -- I think everybody could agree if you, let's say, go on a search engine, you type -- and you live in Houston and you type in World Series and the biased algorithm sends you to the "Nats" site, I think we'd all agree that's a good bias. All right. So let's just get that straight. Anybody watching online from Houston, I'll give you my e-mail.

But, no, seriously, that I, a hundred-percent, agree with Nicol, bad bias, and we should not do that, and there's good bias. Here's an example of good bias: You're running on an HIV campaign online, you have a limited amount of money, you want that bias towards people who are like more likely to

have HIV or HIV risks. So you might not target it towards certain demographics and you target it towards other. That's one.

The second would be where you want to intentionally correct for existing bias in society, so imagine a case of a medical school that gets, you know, thousands of applications for doctors and historically there's been bias against black applicants. You could design an algorithm that says we're going to bias it to correct for that and give black applicants a little higher score, or something like that. So that's an AI algorithm that would be biased, but it would be biased in the sense of trying to get rid of other kinds of biases.

Another would be where an algorithm is biased, but it's less biased than what we have already in society, and a good example of that is facial recognition. Everybody's all up in arms that facial is just teeny, teeny little bit more biased against some groups than others, and to be fair, we should all work to -- people developing those should work against it, but for example, there are studies in looking at lineups and police lineups people who are -- just, in general, only get that right 60 percent of the time and if it's

cross-race, if you're white looking at a black or black looking at a white, you only get it right 45 percent of the time.

Using the NIST studies on the best and effective facial recognition, one-percent error. Now, imagine that that one percent is .12 percent for blacks and .999 for whites, that's biased, but it would lean towards a better outcome because it's less biased than humans. A fourth would be something where you could even have an algorithm that's more biased, but it's still better, and let me give you an example of that: Imagine -- you mentioned X-rays, or a --

MR. WEST: CT.

MR. ATKINSON: -- CT. Imagine an X-ray or a CT scan, you have an AI algorithm for that, and let's say doctors get it right 85 percent of the time. This algorithm gets it right 95 percent of the time for women and 80 percent of the time for men for an overall accuracy rate of 92.5 percent. So in one part, that's bad; it's more biased. Another part, it's less biased, but overall it's less biased. I think that would be a better algorithm because you could say, okay, we know that the bias is not as good for men, therefore you better actually have a human that would look at it again in a different way.

And the last would be where there's -- and, Nicol, where there's little harm. I mean, we just don't care. A good example for that is YouTube is intentionally biased against Scottish people. So if you tried to look at a Scot, I mean, they talk funny. Let's just all acknowledge that, okay. If you put it in a YouTube video, or, you know, that's where the heavy brogue Scottish accent, it gets the transcription right for deaf people 53 percent of the time whereas if you're talking like you're a Midwesterner with perfect diction, it'll get it right a lot of the time.

Does anybody really care? You know, that's a problem.

SPEAKER: I care.

MS. LEE: Right,.

MR. ATKINSON: Where are you from, sir?

MS. LEE: A Scottish person cares, exactly.

MR. ATKINSON: On the other hand, there was a case where an Irish woman who had been -- I forget what it was -- it was some application for residency or something like that and they used her voice and it couldn't recognize her voice, and so she got denied. Well, come on, you can't -- and to get to Nicol's point of it, it depends on the application. If it's an important application with real harm, then bias is super important. If it's a frivolous one, like a dating app, we shouldn't really I think sweat too much about it.

MR. WEST: Okay, so, Nicol, in our AI paper series, we are looking for fresh ideas about AI and new ways of thinking about the topic, so one idea that you have written about is ENERGY STAR rating systems for AI software. So what would that look like and how would that help us deal with AI issues?

MS. LEE: Yeah, so that's interesting. So in having these conversations particularly with Rob and John and others, you know, the question becomes then, you know, what do we do about it, and I'm sure we'll talk about this as we go deeper in the panel; these issues are really hard to regulate and they're hard to legislate, because you have to figure out, you know, one, what's been the collective impact on populations, where in the model is the bias showing up more prominently, who is liable for the bias. Is it the licensed person, is it the developer, you know, and all of that kind of comes into play.

So what I've been trying to figure out is, is there some type of model that introduces

some level of
self-regulatory behavior from companies while at the same time still looking at the studies and standards that NIST provides, but then also integrates customers into this. And so one of the areas, as my refrigerator broke a few months ago, that came to mind was the yellow label that influences me to purchase an appliance, and some of you remember years ago we didn't have this ENERGY STAR rating when it came to appliances.

We essentially went in, we bought a product, that product came in and we found out later through our bills whether it was taking up too much electricity or was a washing machine that was just over abundant with water, or there was other -- you know, the material wasn't environmentally sound, and today with the ENERGY STAR rating, we actually know what goes into that product and we have a sense of its input and its output.

I believe, because in many respects the conversations that we're having about AI and the economy in which it is actually derived from which is our data, it really is dependent upon the trust of the customer to the company with our ability to use that, and so Tom and I were actually just talking about this overseas last week. He's a scientist. There's an extent to which this almost mirrors when we went from HTTP to HTTPS where we had to raise our awareness as consumers as to what is a trustful environment for us to actually operate in.

So what does a ENERGY STAR rating algorithm look like? And this is really early on. We're still sort of going out there and asking Quest Companies questions as well as civil society organizations and academics. It has technical standards that are very much not necessarily where people like myself want to know what the equation is, but I have an understanding by the company that they have tested this in multiple context, so they've not only just used primary data or testing, but they've maybe done secondary or tertiary testing on the model.

It has certain protocol when it comes to the company's ability to de-bias the algorithm or to trigger it to de-bias if it actually hits what John is talking about,
this -- I use the analogy of algorithms like a snake, you know, under the ocean that it just keeps picking up all the sand and over time it doesn't look like a snake. It looks like something else, right. But a really good technical solution allows for some baked-in de-biasing that allows a company to know if something

has gone wrong or awry with that algorithm.

The sample is representative, and if it's not, consumers know about it. So in facial recognition technology, for example, it has a hard time already recognizing people of darker skin complexions. I change my hair very often. If I relied upon facial recognition to do certain things it would not recognize me because it has a worst time, and this is a fact, identifying African American woman with changed hairstyles. So I either want to know that in some respect as part of this ENERGY STAR rating or I want to know that a company has done their best to increase the sample, that it's representative of different scenarios. So that's the first thing.

I'll be quick. An ENERGY STAR rating, and again thinking about my washing machine, has a presumption of where there may be, you know, barriers or gap stops of things it cannot do. When we take certain medications, it says, "Don't take this and go to New Orleans and start drinking," right? "Make sure you eat this with food." And I think that there is some level of responsibility of any type of algorithm, and this is not necessarily something as Rob said, that I'm going to look for with my shoes. But if I'm using a credit application algorithm or an employment application algorithm, I like to see some transparency around that. What is this algorithm not going to be able to do that I should know about?

It has also the third part, consumer feedback. We know the online rating system in this country when it comes to online services and products is real. I don't know about you, but any hotel that I stay at, I look at the review. In this ENERGY STAR rating, it allows consumers the opportunity to see how this algorithm played out in other context or the experiences of people.

And then finally, it's an opportunity for businesses and industry to work together on best practices, and, you know, NIST been really good at actually bringing companies together with academics, et cetera, but the biggest challenge we face right now in this industry is that people like me who are just, you know, used to being the sociology major that nobody wanted to talk to because they didn't know what we were going to do, including my mother, didn't know that the value that we had was the social science background to sort of match with the technical side of it.

To give you some idea that, for example, the use of ZIP code or census track that primarily indicates what places you don't want to find rich people may lend itself to the exclusion of low-income areas. That actually happened, and one of the examples when Amazon was first developing its

Prime, they actually put in these online proxies, ZIP code, income, and they excluded black, poor neighborhoods from being Prime customers.

We need, again, in this final piece the extent to which the company can say we have these cross-pollinated workgroups. We have engineers and academics that work on that. Now, some would say that would be a burdensome process to any company to take on in terms of this label, and I would say they're wrong because the online economy is dependent on our trust. And we've seen these trends, much like we have seen in the automobile industry where that trust pushes us as consumers to want to engage, you know, our data, our time, our money in certain things.

So it's a product, Darrell, that we're working on here. It's tough because as the engineers will tell me, I don't know if I can do that. I suggest that they can. If I have anything to do with it, we're going to actually do something like that. But this rating service as standard kind of moves away from what we've heard so far which I think, and I'll put this out there, are these models legislators don't know how to read explainibility models when it comes to algorithms.

If you're not a data science, you don't know what's under the hood. The whole idea that a legislator is going to be able to tell you what the black box said to get to the outcome is totally ridiculous. So this at least allows for I think cooperative, collaborative process to get consumers at a place where they have some literacy.

MR. WEST: Okay. Nicol, don't worry. We value having sociologists on the staff (inaudible) of --

MS. LEE: Thank you.

MR. WEST: We want to understand the societal ramifications of this technology.

MS. LEE: I know my mother would be proud. She didn't know what I was going to do for a long time.

MR. WEST: Now she knows, so. John, you have written about deep-fake videos which are edits or manipulation designed to harm specific people, do we need new rules in this area?

MR. VILLASENOR: Deepfakes is such a complicated question. So deepfakes are AI-driven modifications, so they are used to make a person appear to say and/or do something they didn't say or do. I guess one initial thing I'll say is they're not necessarily bad. I'll give an example. There is a

Salvador Dali museum in Florida that has a

deepfake of Salvador Dali. When the visitors to the museum walk in, they can interact with this deepfake, and so just the fact that something is a deepfake doesn't inherently mean that it's negative, but a lot of the uses have been negative.

The most common uses have been for pornographic purposes. I know there is placing people in pornographic scenes where they never actually were, and, of course, in the political context as well. So the legislation question is interesting because first of all there are a number of frameworks, and I'm not saying they're necessarily sufficient, but I'm also not saying they're necessarily insufficient. There's a number of frameworks that are out there already. There is copyright law. There is torrents like intentional infliction of emotional distress. There is the right of publicity.

And then you also have to sort of address, you know, be mindful of things like the First Amendment, obviously, and CDA230 which is, you know, the framework that thought of online services, you know, that protects them to some extent from liability for what other people put up there. I'll give you an example of one of the challenges here: California, where I live, recently enacted two new laws aimed at the

deepfakes, one targeting a pornographic deepfakes and the other targeting election, or political deepfakes, but the political one, what it basically says is you can't really distribute a deepfake unless it's marked as such 60 days or less before an election and so when I see that I think, well, you know, does that mean if you do it 65 days or 70 or 80 days before an election.

I'm not sure how effective it's going to be because, you know, somebody who has a nefarious intent can easily release something outside the time window contemplated by that law knowing that it was going to be spread on the wild, you know, so that's one concern. Another concern is, you know, it's easy to write laws that either create civil causes of action or criminalize or both deepfakes. It's harder to write them in a way that doesn't inadvertently sweep in things like parody, right?

MS. LEE: Mm-hmm.

MR. VILLASENOR: And so, you know, you risk this situation where you do achieve your goal of making it harder from a legal standpoint to giving people more tools to fight deepfakes, but in doing so you might find that, you know, *Saturday Night Live*, 10 years from now, if they're doing

deepfake technologies to create, you know, parodies of political figures, you know, runs afoul of, you know, some law that inadvertently catches them. And so it's a complicated thing to legislate with sort of sufficient precision to not accidentally create collateral damage for what we would all agree is protected expression.

MR. WEST: So, Rob, you have an interesting new paper that notes the changing political environment in regard to the technology sector and the paper notes you're worried that techlash may lead to counterproductive policy restriction, so I'm just wondering, can you give us examples of that policy as based on people's concerns about technology that we should want to avoid?

MR. ATKINSON: Yeah, it's almost harder to do the opposite, you know, good policies. You know, we released a report on Monday called the "Policymaker's Guide to the Techlash Advertisement", ITF.org. You can download it. You know, Nicol mentioned AI and education, and right now the Chinese government is putting in, you know, billions of dollars to transform their education system around personalized, customized learning because they know that's the best way to get their kids educated, and also around using AI so the teachers don't have to do all this mundane paperwork and all this stuff.

I mean, it's pretty transformative. Yet in the U.S., virtually every effort to use AI in education has been opposed, and it's been opposed either for privacy reasons or bias reasons, and that's a good example of that. Now, can you do AI in education that doesn't have bias? Absolutely, you can. And I know Nicol's not saying this, but there are some people who say that AI is -- I forget who said it recently --

MS. LEE: Right. So some people say it's -- yeah.

MR. ATKINSON: A prominent person said this recently, "Because American society is inherently biased, there is simply no way to have AI that's not inherently biased." First of all, that's not -- again, as I said before, it's not the right metric. The right metric is at less biased, but number one, I just don’t fundamentally believe that. I think you can do that. It's a little --

MS. LEE: Ahhh. I'll respond afterwards.

MR. ATKINSON: You know, another thing about that. If you've got time to think about AI as this sort of magic pixie dust, you know, that it's just all powerful, it's just a set of code, and it can do

what we want it to do, and I think people, that you know, they overstate the power of it. I encourage you to read Gary Marcus' new book on AI. It's just fantastic; came out about a month ago. He's an AI scientist from NYU.

But so that would be a good example. Another example, I think there are some people who have proposed a national AI regulator. Now, imagine that we were in 1994 and the browser was coming out and the Worldwide Web, and we say, you know what? We need an Internet regulator. Well, thankfully we didn't do that, and Ira Magaziner was advising Bill Clinton to do a light touch one. The idea of an Internet regulator would have made no sense whatsoever because the Internet is everything. It became everything. AI will become everything. It will be used in agriculture; it is being used in all these other areas, so and even having an Uber regulator makes no sense.

I agree with John's point, is that we're going to have an existing set of laws and legal traditions and background. We're going to have existing agencies, so if you try to do bad AI in securities trading, the SEC is going to go after you. If you try to do it in housing, HUD and other housing agencies will go after you. So I worry that we're really backing ourselves very quickly into an AI techlash where we see it all as a problem.

You know, for every hundred really good effective non-biased AI rollouts, there may be one -- I mean, who knows what these numbers, but, you know, maybe one that's not good. The one is the one that will be on the front page of the *New York Times* and the 99 you'll never hear about. So I think we have to keep that in mind. There's lots and lots of people out there working on this challenge and it's super important that there are advocacy groups and think-tanks that are holding people accountable; don't get me wrong on that. But I worry, though, that we're just going to say, no, no more AI. It's just too risky.

MS. LEE: Yeah. Can I --

MR. WEST: Okay, Nicol, you wanted to respond to that?

MS. LEE: So I agree with Rob. I think that, you know, placing like additional regulatory burden makes no sense particularly in innovation where the policymaking will always be behind the innovation and you put certain guardrails, you know, much like we're seeing with, in my opinion, facial recognition, and the entire bans, that there are, you know, some areas we need to sort of play some

scenarios out or put guardrails around parts of it as opposed to marginalizing the whole thing.

But let me tell you why I believe that bias, though, is inherently baked into AI systems. I'll just tell you two reasons. One, the people who are developing them are not diverse, so let's just start there. When we actually develop computer bias systems and models, we come with our values, our norms, our assumptions about what the world looks like, and when you come with those values, norms, and assumptions and you work in Silicon Valley which is 98, 97 percent white male, that's what the AI starts looking like and that's why the bias is inherently baked into the system, so what does that look like in search queries.

Search query agrees, and this is all publicly available information, when a person put in a search for happy teenagers on a Google very innocuous search, it delivered results of white teenagers smiling and black teenagers with mugshots. When a person put in a search for African Americans, it delivered primates as opposed to people. And when those developers were asked, "Why did this happen," you know what their response was, "We didn't know that that was going to happen the way the data was tagged." This is all available on the paper.

And so what that tells me -- Amazon did the same thing. A (inaudible) algorithm that was supposed to bring more people to the engineering department, particularly women, based on training data that came from 10 years of resumes at Amazon which were predominantly men. Every woman was knocked out if had a woman's school, Mary, Jane, or whatever the name, because of the way that the algorithm is designed.

So I think we have to solve that problem. We cannot have ubiquitously deployed models that are affecting all areas of our life built by people who are homogenous white men, period. So let me just start there. So I think they kind of bias. I think the second thing is the other challenge that we have is the technical nature of an algorithm particularly when it comes to search prioritization, further creates tribalism and polarization.

And let me tell you what that looks like. So we all -- who's on Facebook? Just raise your hand by a show of hands. It's not an endorsement of the company. I just want to know who's spending their time there, right? If you think about that, some of us are on the slide, some of us are, you know, looking at data that could be like a playground, right? On the

swing, some of us may be playing in the sandbox, but we're all on Facebook.

Well, let me tell you who those people are. Some of us who are in the sandbox are white supremacist. Some of us on the slides are people who like to travel. Some of us in, you know, playing with the monkey wrenches, and we never see each other or talk to each other because the way that the algorithm is designed, it further reinforces the things that you see because it picks up on your behavior. That, in and of itself, I believe creates the areas of the sociologist systems of oppression and domination that come by who actually creates the algorithm.

So I'm not necessarily saying that we can't change that because we can by actually increasing a pipeline of diverse creators and people from different disciplines and creating more interdisciplinary work in this, but computers aren't biased. They just don't wake up and say I'm a bias against people on the computer today. No, (laughs) it's the person behind the program of these models and what they bring in their set of assumptions. So, Rob, I agree with you, but that has consequences --

MR. ATKINSON: Okay.

MS. LEE: Well, I do. I really do, but (laughter) -- okay, I know. I like Rob, though, we're friends. I mean, we -- oh, come here, Rob, (inaudible).

MR. ATKINSON: We don't agree.

MR. LEE: That one percent may seem innocuous, but it's not, and that's what the challenge is, right?

MR. ATKINSON: Can I respond to that, Darrell, real quick?

MR. WEST: You can respond, but this is the first time we've had a group hug at a Brookings' panel, I (inaudible)

MR. ATKINSON: All right, fine. We could all three of us here -- but I don't know you, John, so I don't know that that'd be appropriate. Couple of things: One, and both of these issues are -- one of the 23 techlash myths in our report, one of them is around demographic diversities, Silicon Valley. Actually Silicon Valley companies are more diverse than the population of computer scientists when it comes to gender, so if you just look at the number of women, the share of women who are -- the share of graduates who are women, Silicon Valley companies employ more women in computer science than the

share.

Now, you can say, well, they should employ more, but if they do that then it means that the non-Silicon Valley companies have even fewer women. They are not quite as representative for Hispanics and Blacks, but it's not zero. They're slightly under what the men be (inaudible), and they should do more; I totally agree, and they are doing more. So the idea that somehow there's no diversity in the tech community I think is just -- it's an overstatement. Secondly, I just reject that view. I'm sorry. I think that given the attention -- I think there are risks of developing a biased algorithm is higher if you don't have diversity, but it's not a hundred percent. I just reject the notion that you could have a team of white males who are -- you know, they read all the, you know, woke publications and they're like incredibly sensitive. I think they could design an algorithm that's unbiased. There's nothing inherent about it. It's just it's sort of laziness and sort of ways of thinking, so I just reject that --

MS. LEE: Okay. Well, can I --

MR. ATKINSON: I want to say the last thing, which is search polarization. Again I just don't -- I mean, if you look at the evidence on that, the most polarized people are people who are age 65 or over. Polarization started long before the Internet (inaudible). There's a number of good studies that actually show that social media forces people or encourages people to see more diverse things than they would otherwise, so, again, I don't buy into that. It's not to say it's not a problem, and on the data-set thing, again --

MS. LEE: Don't buy into that, too.

MR. ATKINSON: -- these people have PhDs in data science. I'm telling you there is lots and lots of ways that you can control for certain data sets and adjust for that. It's not like it's inherent that you have to have it.

MR. WEST: A quick response and then I want to hear (inaudible).

MS. LEE: So I promise I'll be really quick and then I'm going to have Rob back on my panel. So I want to reject that as well because I think the challenge that we're having when we look at technology is that we take it outside of the context of the historical systemic and qualities that we've already actually manifested. So I can take your same example around workforce diversity not being a problem and I can look at myself --

MR. ATKINSON: I didn't say this.

MS. LEE: -- well, no, but --

MR. ATKINSON: I didn't say that.

MS. LEE: -- you said in terms of the --

MR. ATKINSON: I said it's --

MS. LEE: -- a limited problem, right?

MR. ATKINSON: It's more diverse than you had said. That's all.

MS. LEE: Okay, so I can say the same thing about higher education, okay, and take it totally out of the technology space, talk about the number of degreed people of color that come out of PhD programs. When I got a PhD, it was less than .005 percent, and a lot of that has a lot to do with the structural inequality that goes into who's actually pursuing a degree in higher education, who gets into the college track, and I share that example because I think in the technology space, we make the assumption that the technology in and of itself is a solution to changing everything that we have historically dealt with over years and somehow the technology is going to make it better. It is a problem.

It's not necessarily a bad thing if five very woke white men sit in a room (laughs) and come up with a really good algorithm. But if that algorithm is committed to testing the pain thresholds for women and they're not at the table, then that's even more problematic because we're putting out a product that's not representative of that. And because the tech space, and I'll finish here, Darrell, is really about a race to market for PhDs for us. We cannot do any studies that impact human people without a comprehensive internal review board process. We have to insure that children and other people that we're studying are not hurt.

In the tech space, if I work in a tech company, I got to beat the next person to get that product out whether or not that sample is weighted in terms of, you know, regression analysis has a certain mean where it's got a representative population of women, people of color, people over 50, and, to me, that’s a challenge. Like, I am in no way suggesting that we need to put more policy guardrails. What I suggest is that we need to have more conversation that these are inherently happening in these processes that are not dictated by levels of fairness, ethics, and equality. That's all I got to say there.

MR. WEST: Okay, John, quick response.

MR. VILLASENOR: I just wanted to add just another --

MS. LEE: I say, poor John.

MR. VILLASENOR: No, no, obviously, (inaudible) different to illustrate some of the complexities of the bias question: Risk assessment in criminal justice. So it turns out, and this won't surprise anybody, we talk about gender, and men reoffend at a far higher rate than women. The recidivism rates for men; it's particularly for violent crimes are much higher. So you have a really interesting question. When you're using risk assessment algorithm, do you consider gender?

Because the argument against considering gender is that gender is a protected characteristic and, you know, there's laws about that, but the argument in favor of considering gender is because it is prejudicial against women to basically evaluate their risk in a pool that includes this much higher risk pool of men, and so I illustrate that just that some of these questions don't have really clean answers, because a reasonable person could make either argument about whether or not to include gender in recidivism risk assessments. I just wanted to make that point.

MR. WEST: Okay, so I have a last question for the panel, then we're going to open the floor to questions from the audience, and the question is: So we've outlined a number of different AI problems; where should we look to fix some of these issues, companies versus government? So what responsibilities should companies have to fix AI problems, so should they have ethics review boards to look at AI deployments and their effects on various parts of the population; should there be third-party audits of what they're doing, and then what should the role of government be?

So, for example, there are some local governments that have enacted outright bans on the use of facial recognition software. Some people, including myself, suggest the need for time limits on data retention. What should this mix be of company responsibility versus government responsibility? John, why don't we start with you.

MR. VILLASENOR: That's a hard question. You know, I guess I'd advocate for balance in the sense that, you know, echoing somewhat what Nicol said is that, you know, you turn the dial too far to the right in terms of, you know, oppressive oversight, you kill innovation. On the other hand, completely unregulated, you know, untampered work in this area, you know, could lead to some very

significant harm, so I would like to think that we can strive for balance. I also think just to -- you know, one thing I'd say is that the landscape looks very different for smaller companies than for bigger companies.

You know, for all the criticisms a company like Facebook or Amazon or Microsoft might receive, they also have the resources when these problems are pointed out to then put people on them, whereas you look at a very small company that's just trying to get a product out to market, they might not feel that they can put the resources in to preemptively address some of these, you know, more negative consequences that can occur. So I think we need to sort of think about the different -- not only think about companies, but within that eco system who are the different kinds of companies and what might their particular incentives be to sort of, you know, insure, you know, integrity in the algorithms are not.

MR. WEST: Rob, your thoughts on that?

MR. ATKINSON: You know, I mean, I would be -- tend to -- you know, look it can't be either/or, and you're either a state (inaudible) libertarian, and neither of those positions are wrong, but I think the most important thing to understand are this is really, really new. I mean, this is early on, you know, IEE (phonetic) developing its ethics (inaudible). You have academic researchers funded by NIH to look at these questions around from the AI science component. You have people being taught AI ethics in colleges. You've got lots of articles, studies, all this. I think, number one, we should just slow down a bit and calm down a bit.

MS. LEE: Right. (laughs)

MR. ATKINSON: -- and rushing to some sort of, you know, ban, to me, is the worst possible thing we could do. Having said that, and I agree with John that, you know, companies do need to be liable. We wrote a report last year where we called for an algorithm accountability framework. We really don't believe in explainability. You can't explain these things or transparency which has IP implication, but you should be accountable for the algorithm you come out with, and there should be penalties for saying something and doing something else, but I also think we just need to slow down a bit and look, sort of see how things evolve somewhat.

MR. WEST: Nicol?

MS. LEE: And I agree with what John and Rob actually said. I think on the company

side, as I've suggested, I think engaging companies more, you know, in best practice conversation would be really helpful. You know, what are companies doing well, what are companies not doing well, where are there some areas we can pull training data. You know, is there something that we can do together to actually solve this for self-regulatory, you know, what all Darrell talked about, audits, impact statements, all those things in a bucket actually work well, and they work better, I think, when companies come together to speak about it.

I think from a government perspective, because this is such a complicated issue, I do think that government could do one thing and just one thing to at least start the ball which is to remind tech companies and this goes with what I've said all morning, that it is against the law to discriminate against people online. That we have the Fair Credit Reporting Act -- we have the Fair Credit Act, the Employment Act, the Housing Act, we have other civil rights statutes that were prior to the Internet that should be legally recognizable on the online space. That's a real simple, simple fix for the government to actually come in and do without finding themselves inserted in areas in which they have very little knowledge at this point.

MR. WEST: Okay, let's open the floor to questions from the audience. We have people with microphones, so right here is a gentleman with a question. If you can give us your name and organization and if you can keep your question short, so we can get to more people.

MR. ENACHIA: Thank you very much. International Urban Alliance. My name is Enachia. I congratulate you. It's fascinating subject. My question is, have you ever attempted to connect this AI problems and litigation to group dynamics, hurting mentality and behavior, and more generally human nature and its biases, because some authors believe we are already cyborgs. We're not.

MR. WEST: Any thoughts? Okay, I guess the answer is no.

MS. LEE: Yeah. No. Yeah, I mean, I would just say the same thing that we're talking about. We're seeing more attempts where people are being placed within communities, in my opinion, where those behaviors are amplified that do have the ability to repress, you know -- there was a study I was just reading about, happiness and other things, so I think, again, because we live now in an always on, always connected society -- as a sociologist, one of the greatest books I read was Robert Putnam's *Bowling Alone*, which was around the breakup of civic engagement, and the fact that people don't get

together. It has the potential, particularly when we look at deepfakes and other neural manipulations, to actually go further.

MR. ENACHIA: Thank you.

MR. WEST: Okay. Right here. This is a gentleman with a question.

MR. APGAR: Thank you. Sandy Apgar, CSIS. One of you mentioned the principle that companies would be responsible for the algorithms they create.

MR. ATKINSON: I think that was me. Yeah.

MR. APGAR: And I've -- as a non-lawyer, I don't understand in a creative industry, or profession creating algorithms as distinct from writing technical algorithms, how a company or university or any other institution should be ethically responsible for the individuals' or teams' products. Perhaps that's well established law, but as a matter of management and operations, I don't grasp and wonder how you see that principle being established.

MR. ATKINSON: It's well-established already, even independent of AI, that a company is responsible for its products.

MR. APGAR: Products, yes, but this is -- and how is an algorithm which is a creative --

MR. ATKINSON: Yeah, it's a --

MR. APGAR: How is that a product?

MR. ATKINSON: And by the way, algorithms -- and just a taxonomy, algorithms is a broad thing and AI is a subset, and so algorithms long predate AI and so companies, for example, that pre-AI had algorithms for making auto -- you know, helping them with loan decisions and these loan decisions were discriminatory. They were responsible for those algorithms even though you might argue that it was a creative act to create them. So the fact that something's creative -- and I think code writing is creative. Algorithm creation is creative, but that in no way relieves the creators of responsibility for what they've let loose into the marketplace, because remember they're doing so, so they can receive the economic benefits --

MS. LEE: Right.

MR. ATKINSON: -- from those things, and part of that bargain is that, you know, you get the money, but you also get the responsibility if things go wrong.

MS. LEE: Right. If I can --

MR. WEST: If I can tag onto that. I had a paper that came out a year ago on kind of the responsibilities that companies do have in the AI space, and I want to point out most companies actually do accept responsibility. Companies are starting to develop ethical principles. Some companies have established ethics review boards so that they can look at the AI algorithm early in the development and then also look at the post-deployment stage in terms of the actual outcomes, how it's affecting people, how's it affecting various protected groups, so I think that there are lots of ways that companies do need to step up and some of them actually are stepping up.

MS. LEE: Yeah, and I'll just to say real quickly, so with the health care algorithm, it's -- and actually was noted in the *New York Times* article that it was not known who actually licensed that algorithm, and if you go onto the second article, what basically happened was United Healthcare, apparently, paid the company to create the algorithm who is now taking some responsibility for the discriminatory effect. So I think going back, you know, we sometimes think that this is somewhere in the air, there are companies that are basically working for big companies or departments within companies where in our paper we actually attributed to -- I think Rob uses, too -- operators. You license, you diffuse. You know, it's just not the algorithm by itself, which is liable, somebody's actually making it.

MR. ATKINSON: I'll just say it quickly. I think one of the debates is going to be is that the algorithm builder or the algorithm --

MS. LEE: Right.

MR. ATKINSON: -- operator, and I would argue it should be the operator. If the operator gets sued, they can sue the developer, but I think fundamentally if you got hurt by an algorithm from your health insurance company, you sue your health insurance company; you don't sue --

MS. LEE: The creator.

MR. ATKINSON: -- the builder of it.

MR. WEST: Okay, in the very back on the aisle, there's a woman -- right there. Yeah. Give us your name and organization.

DR. AMUSVIE: My name is Dr. Fatina Amusvie. I have PhD in AI, and I believe in all

what you're saying because I think the company, at the moment, they are at initial stages of the AI. Everybody is learning about AI, everybody learning about the biased and unbiased of the AI, so what happened, the company goes ahead and just like create the model and then present it, but in fact they should study it more because if you add more valuables or more data to the model and then you train it and test it, you will have different outcome, so what happened, I feel like, you know, they should not go ahead right away and just present their output because it's not fully hundred percent according to the data valuable, what they have, because they should add more data to the system, like, for example, washing machine.

The washing machine before like 20 years ago, it was not like, you know, automated and (inaudible), the same things as AI, and I believe as Dr. John, that they should be a little bit more patient, not running like a horse. They should hold their horses and slowly, slowly learning about AI and that will be --

MR. WEST: Thank you for your comment. There's a question up here. Right there. Yeah.

MS. BARTON: Hi, Genie Barton and I was privileged enough to be --

MR. WEST: If you hold the little -- yeah, okay.

MS. BARTON: Yeah, I -- privileged enough to be Nicol's coauthor on her original paper. My question is, it seemed -- I agree with many of the things said by everyone up there. I think Darrell said something very valuable about testing which we need to be sure when we have an algorithm that we do testing before, middle, and look at whether the assumptions underlying the correlations that we're making are actually the best assumptions. That was what was wrong in the health care example that Nicol gave, is the correlation was amount spent on health care deciding how sick the person was rather than looking at their actual health. And because Blacks spend less on health care, the model was erroneous.

When it was tested appropriately, the error was found. So I agree that it would be good to have product liability, but we need to develop standards that are applicable for how do we build the algorithm, what is the degree of confidence that we need depending on the situation, and I think that the best practice model and the caution that Rob has suggested are really important and also let's enforce the existing laws we have with respect to really important decisions about people's lives. So I'd just like to

ask everyone, how do you think we can get people to understand what confidence levels are need and how to develop, you know, that little thing you want, Nicol --

MS. LEE: A ENERGY STAR rating?

MS. BARTON: Yeah. (laughs) Of really what's been the testing, what's the confidence level, is there ongoing testing?

MS. LEE: Right. I'll --

MR. WEST: Okay, any responses?

MS. LEE: -- quickly just respond to that, too. I mean, I think the confidence level is really interesting because there are some companies, and we've seen this actually with facial recognition, and some of the examples Rob gave as well, where companies will say, "In our lab, the confidence level was X, and then it gets deployed by a third-party actor, like we've actually seen with facial recognition by law enforcement where they're not really paying attention to confidence levels or researchers who actually partition that data and they're able to put it in other context, then finding the confidence level threshold is not there.

This is why I think John's conversation about product liability is really important, right? Because in a certain respect, companies are indemnifying them self from this because they feel that if you use X, Y, and Z and it generates this confidence, then it's fine. But as we've discussed today, this -- you know, facial recognition wasn't thought to be used in ways that we're seeing it now. I mean, if you look at passport identification, deportation, that those use cases trigger protected classes and other categories which make, I think, Genie, to your question this conversation about testing really important.

I would just say to the wait up -- you know, hold on a minute, hold up, wait-a-minute kind of analogy with innovation, the only challenge with that is that innovation just goes so fast, and the benefits of AI are also very vast and very good when it comes to health care and breast cancer diagnosis, and the climate change and environment, and energy. So, again, it kind of goes back to I think what my colleagues are saying, where's the balance, so that you can also have somebody's really interesting innovations that surpass human understanding while at the same time, you know, employ certain areas where you have good governance and liability around it.

MR. WEST: Okay, I think we have time just for one more question. Right here, and

there's a microphone right behind you.

MS. MAYLONE: Thank you. My name is Connie Maylone. I am a New York state union of teachers, okay, high school teacher, 32 years; I do not know how to write code. But I do know how to teach, and the one thing is to identify a goal, so what is the overarching goal of any particular algorithm, then which subgroups are your target audiences, what are your target audiences within your goal. Identify your target audiences, write many algorithms which would identify, which would apply to each target audience, and then interface those many algorithms rather than writing just one overarching algorithm or a lesson plan.

We don't do that anymore in school. We don't write an overarching lesson plan. We have heterogeneous groups of kids and we mini-chunk the lesson plan. Why can't that be done with an algorithm? I understand you need an overriding goal, have an overriding goal, identify your subgroups, many algories (phonetic), if that's a word, many --

MR. WEST: That's a new word.

MS. LEE: That's a good word.

MS. MAYLONE: Okay. Many algories, the different subgroups, interface those many algorithms.

MR. WEST: Responses? Rob?

MR. ATKINSON: Well, I would encourage you to read the best book on AI, which is by Pedro Domingos who's at University of Washington, CS Department, called *The Master Algorithm*, and if you read that book, you'll get the answer, which is you can't do what you're saying. You got to create a sort of overarching algorithm. But I want to come back to this confidence level thing because I think there's so much misunderstanding about that.

It's if we assume -- I mean, at one level a lot of the people will rightly will say, you know, American society is embedded with bias, but then you go, what's the standard for AI? Well, the standard for AI is that every other decision is completely fair, and that is the wrong standard. The right standard is, is AI better? So imagine an AI algorithm that discriminates against women, but it only discriminates one percent whereas -- I'm just making the numbers up --

MS. LEE: There you go, that one percent.

MR. ATKINSON: -- but the male, you know, HR people discriminate at 10 percent. That's really bad. I would much rather have, and I'm a man here, but --

MS. LEE: I know, right.

MR. ATKINSON: -- (inaudible), the AI algorithm, so that's point number one. Point number two is this confidence level thing, and that to me is where there's again so much misunderstanding, and one of the groups who's done the most harm to this is the ACLU. So the ACLU reported on facial analysis, by the way, not facial recognition. Facial analysis is super hard. It would take a look at me and would go, how old am I, and if you've seen that app on your phone, yeah, that's facial analysis.

Facial recognition is: It knows exactly my measurements and those are super accurate. But in any case, what it did is it used facial analysis to analyze members of Congress or other things, and it used a 70-percent confidence interval. So, yeah, they're going to be an enormous number of mistakes, and yet the developers of these algorithms say you should never use them unless it's like a two-percent error rate, and the ACLU ignored that, come out with a study that says these are inherently biased, they make all these mistakes. Of course they do when you dial down the confidence interval's so low.

So then what did they do? All this came out, we reported on it, a couple of other people; they do another study doing the exact same thing, and that study gets widely picked up by the *New York Times*, all the media who are not computer scientists or stats people, and it said, "Oh, look, once again, these algorithms are biased." And again, it really goes, I think, to the point of, you really have to look at this and you have to look at the motivation. The motivation of the ACLU is to not have any facial recognition technology in the world. That's their motivation.

So you can -- if you use the right confidence intervals, you can be pretty sure that they're going to work well, and now, again, part of it is, are the people who are using it, are they trained, do they know don't turn that dial down, keep it at two percent or whatever, and that's an issue but when you intentionally turn the dial down to get biased results, to me that's not fair.

MS. LEE: Okay, so just real quick. I promise, Darrell. So the thing with -- I'm not commenting on the ACLU on that, but what I do know about facial recognition is two things. One, the study that they did talk about, particularly the one of the congressional members, said congressional

black caucus members as criminals --

MR. ATKINSON: Because they used --

MS. LEE: -- because they use a --

MR. ATKINSON: -- the 70-percent --

MS. LEE: Because they used 70-percent --

MR. ATKINSON: -- and they didn't --

MS. LEE: I'm going somewhere with this.

MR. ATKINSON: All right. If you use one percent, you don't get that error.

MS. LEE: How many of you in this room are engineers? Just raise your hand. Okay. The majority of us are not engineers in society and we therefore will not use the confidence level that a company actually says. We will err on the side of the 70 percent simply because the context in which we're deploying the particular technology suits whatever interest or outcome or a goal or a lesson plan that we have. We just need to get better with letting technologists know that where these models are being deployed are not within their learning lab; that they actually apply to real people on credit situations, employment applications.

So, Rob, to your point, I completely agree with it, what you're saying that the confidence level matters, but you cannot as a company sort of say if you didn't use a two-percent confidence level and more people got incarcerated or over-criminalized as a result of a 70 percent or congressional black caucus members came back with their mugshots because they all committed a crime, that's not acceptable, and so either you have to do a better education on the confidence level affecting the results or you have to design the product, just like my ENERGY STAR standard, with where this product going to have limitations based on how you use it. So that --

MR. ATKINSON: So if you look at --

MS. LEE: Oh, man. Darrell's like, we're just going overtime.

MR. ATKINSON: If you look at police use of these, they use the right confidence interval in 99.9 percent of the cases in this country. What I worry about in the confidence interval are advocacy groups who are trying to manipulate public opinion for a particular cause. I don't worry about the police. I don't worry about TSA. You got -- talk to

TSA --

MS. LEE: But I do.

MR. ATKINSON: I don't --

MS. LEE: -- I worry about the police because they look at me first. No, no, I think we should -- -- I worry with the police. You may not worry about it, but I sure have to worry about it.

MR. WEST: Okay, we'll let the two of you continue this conversation afterwards, but I want to recommend our paper series. Check it out at Brookings.edu, and want to thank John, Rob, and Nicol. And thank you very much for coming. (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