THE BROOKINGS INSTITUTION

WEBINAR

WHAT EXACTLY IS 'RESPONSIBLE AI' IN PRINCIPLE AND IN PRACTICE?

Washington, D.C.

Monday, May 10, 2021

PARTICIPANTS:

NICOL TURNER LEE, Moderator
Senior Fellow and Director, Center for Technology Innovation
The Brookings Institution

NATASHA CRAMPTON
Chief Responsible AI Officer
Microsoft

WILL HURD
Former U.S. Representative (R-Texas)

JULIA STOYANOVICH
Assistant Professor
NYU Tandon School of Engineering

* * * * *

P R O C E E D I N G S

MS. TURNER LEE: Good afternoon, welcome. I'm Dr. Nicol Turner Lee. I am senior fellow in Governance Studies and the director of the Center for Technology Innovation at the Brookings Institution.

I'm glad that you all could join us today and particularly for a topic that I think all of us have top of mind, particularly those of us that watch the artificial intelligence trends. And the conversation we're going to have today is around responsible AI. What exactly is it

Let me tell you why that's important to the work that I do at Brookings. At Brookings, I work on a range of telecommunications issues, as well as tech policy issues, and I hold a portfolio in artificial intelligence. And in the space of artificial intelligence I'm particularly interested in algorithmic bias.

So when I think of responsible AI, I think of these words, like, ethical. I think of the words, like, transparent and fairness. I know I have added lawful to it, and there are other folks who have put among the umbrella of responsible AI a lot of different factors.

But I'm a sociologist and so whatever I think about responsible AI is often, you know, incurred because of my background academically. So I'm really excited about this conversation because you all who follow my work and have been looking at what I am trying to do with this energy star rating know that I'm all about interdisciplinary conversation.

So, today, we're going to have a conversation with folks that I think represent three distinct verticals in the responsible AI space. We're joined today by Natasha Crampton, who is the chief responsible AI officer at Microsoft Corporation; we are also are joined by Will Hurd, who is a former member of Congress; and Julia Stoyanovich, who is the assistant professor at the New York University Tandon School of Engineering.

And, Julia, I apologize. I'm a New Yorker, as well, and sometimes my r's, and my o's, and my ands get all mixed up. But I want to thank all of you for joining us today. So we're going to jump right into this conversation.

Let me remind those of you, who are listening, to please submit your questions to at events@brookings.edu, and on the #AIBias. Let's jump right into this conversation, if you all don't mind.

Welcome, welcome.

Natasha, I want to start with you. And I was to start with you because I think you can help us from the industry perspective on framing what is responsible AI? So, first, tell us what you're tasked with because I have been around for a long time. I have not heard of a chief responsible AI officer, so would love to hear a little bit more about what you do, and then I'd like to hear what you think of when you think about this concept of responsible AI.

Thanks for joining us, too, Natasha.

MS. CRAMPTON: Thanks, Nicol. My job, in a nutshell, is really to put Microsoft's AI principles to work across the company. We assessed our principles back in 2018, to really see the North Star for efforts and they're aligned with the many sets of principles that have been issued since then. So we are focused on seeing that some reliability and safety, privacy and security, inclusiveness, transparency and accountability.

So since 2019, when our Office of Responsible AI was established, I have been really working with colleagues across the company to operationalize those principles, which is really to say that I have been working to define what it means in practice to uphold our principles.

Much of our effort has really been focused on four big building blocks. First, the governance structure to enable progress and accountability, rules to standardize our responsible AI requirements, training and practices to promote a human seeded mindset and tools and processes for implementation.

Our responsible AI governance model really borrows the hub and spoke approach that has worked well across the company to integrate privacy and security and accessibility into our products and services.

So, in our hub, we have three groups: Aether, whose research lead working groups help us understand the state of the art with respect to our principles; the Office of Responsible AI, which I lead, which sets out policies and our governance processes and coordinates the effort across the company; and a third group, which is called, RAISE, Responsible AI Strategy in Engineering, which is really helping our engineering teams to implement responsible AI practices through systems and tools.

And so these three groups work together to set a consistent bar for responsible AI across

the company, and we empower the spokes in our engineering and sales teams to drive initiatives and be accountable for them.

Another part of my job is to actually write the roles to enact our principles. And we do this by a document that we call, Our Responsible AI Standard. We published the first version of this in 2019, really with an eye to learning and a very humble recognition that we were at the beginning of our journey to move from principles to practices.

Our pilot teams really appreciated the examples of how the responsible AI concerns could materialize, but they struggled with open-endedness of the considerations and they were really thirsty for more concrete practices that they could follow.

So towards the end of last year, we began previewing a second version of Our Responsible AI Standard with teams, and this time we have asked teams to work towards requirements that led up to goals. And we know that our engineers are problem solvers, so we really think this goal-driven approach is going to help engage their problem solving incidents.

So with that little bit of flavor for the sorts of things we have been doing internally, I'm going to turn to the second part of Nicol's question which is, you know, why is responsible AI such a focus and why are the roles such as mine starting to emerge?

And, for me, I think it really boils down to two reasons: First, it is completely clear that AI is one of the most transformational technologies of our time. We know that it has the potential to help us with some of the world's biggest challenges, but it cannot achieve that potential if people do not trust it. And to build that trust we have to really proactively lay the groundwork to secure its benefits and anticipate and guard against where it might go wrong as much as possible.

And the second reason I think these sorts of problems decided to emerge is that the types of issues we need to consider with AI systems are quite fundamentally different to the issues we considered when we were doing more traditional forms of software development.

We are asking our teams to think about their systems, not just in a purely technical sense, but in a sociotechnical sense, and so that requires a new muscle. It requires new processes and practices and tools to help guide this work. So I do think with those two fences combined, you're going to continue to see roles, such as mine, crop up in organizations and they're going to share this common

thread as helping to build a culture and a practice of responsible AI.

MS. TURNER LEE: Yeah. I mean I think that's so interesting, right, because I think we saw a similar trend with privacy. For a long time, we didn't see Chief Privacy Officers and now we're seeing Chief Responsible AI Officers. But I think you -- and I want to talk about this a little bit more as we get through the conversation.

I think you layout what some of us, you know, we have more of the definitional side of it, but you laid out the governance side. So how do you take these principles into action? And it sounds like that's from some structural changes, collaborative workmanship, as well as some cultural values, which I think is interest, because in the work that I have done you see a lot of companies often put together piecemeal strategies to figure this out. So I want to come back to this governance side shortly.

I know you don't want me to call you, congressman, but I am for the sake of just the beginning. Congressman Hurd, you have been active at this. I mean before you left office, I had the great, humble privilege of working with you on the Bipartisan Leadership Council on the AI principles. And the BPC really took a look at that with you and Congressman Kelly, as you started to think of what does AI look like on the public policy side or the government side?

Share with me, you know, what you're thinking about when you see this term of responsible AI, and how you have actually taken that, because before you left you also put out some legislation to sort of think about responsible AI from the government perspective.

MR. HURD: Sure, Nicol. It's great to be on, work with you, and such a smart group of people. So, I apologize for the ham-handedness of some of my answers because I am nowhere near as intelligent as the rest of y'all.

You know, this was not something I thought I would be involved in, right. My degree is in computer science, yes. I thought I was going to be a software engineer, and then I got exposed and decided to go, you know, into the CIA and recruited spies and stole secrets all over the world. It was a great job.

And I got frustrated with Congress and decided to run and I got put on a committee because I was the only member of Congress that had a computer science degree and helped build a cybersecurity company and so I kind of got dragged into this. And what I have come to learn, and

especially through this work with the Bipartisan Policy Center, and my good friend, Robin Kelley, to me responsible AI that what we're trying to get to is artificial intelligence that follows the law, period, right.

To me, it's that simple, and we already have rules and regulations on the book. I have learned this from you, Nicol. And so, we already have these rules and regulations
So when it comes from, for a digital tool that interacts with the physical world we have laws.

So let's just follow those laws. And if there is a unique scenario in which it's something that's purely digital -- and I don't know many examples -- I don't have any on the top of my head -- that maybe we may need new laws for that. But let's make sure the tool itself follows the law, and then the application of the tool follows the law.

So, for me, that's responsible AI, and then the future is that responsible AI just becomes AI because the term means it's going to be following the law and doing things that are right and proper. Now that's a simple explanation, but how do we get to that point?

That's the difficult part where we need the smart engineers that are looking at the data and the rules within the algorithms and how you train these things. And so, for me, that's -- and guess what, Congress is actually, you know, built to answer some of these questions, right. And so the role of our elected officials is making sure that our regulatory agencies are applying the existing rules and laws to their use of AI, that they're protecting data the right way. So that's how I look at this perspective and come to these conclusions from my time in Congress.

MS. TURNER LEE: Yeah, you know, look, you know, what I feel about that, and I think most of the people who have been watching my work. I think that there is, like Natasha said, there is governance around it. There is obviously this ethical nature of the AI, the transparency of the AI.

But, my friends, it has to be lawful, right. We cannot have these mistakes. And this is something that Will and I talked a lot about which is, how do you make the new digital world compliant with anti-discrimination?

And if you can't, let's figure out how to do that because people should not be denied housing, you know, loan, or any other critical service because the algorithm or the AI in itself violates what's already been litigated by groups.

So, you know, I'm going to come back to that, right, because I think that is such an

important part of this conversation, but let me go to Julia first. Because, Julia, I know that you're like tiptoeing between the two of these spaces, right. Your work at NYU is all about democratizing AI. I have been just fascinated by what you all have been doing both at the local level, but then also at the academic level. Where do academics fall with this? Because we've heard industry, we've heard government, and now where is the academy when it comes to responsible AI and unpacking this?

MS. STOYANOVICH**:** Thank you for the question, Nicol, and it's such a pleasure to be on this panel. So, to start, I actually just want to say that we all, I think, by definition are outside our comfort zone when we talk about responsible AI. It's just such a heavy topic.

It's AI technology, right, something that this is very, very difficult for us to become comfortable with, even for the technologists among us. But then there is also the responsibility and the ethics. So I don't think that there exists a person in the world today who would say, I am perfectly comfortable with the kind of representation in responsible AI. And, in fact, I'm an expert, right.

So I think that for us to become sufficiently comfortable with this topic, to embrace it as people, to be able to make progress in this really, really difficult space, there are a couple of things that we need to do.

And one of these -- and I hope that we will be able to do this on this panel today already, is to disagree, and to fight, and to have very, you know, strongly worded and enthusiastic conversations about this and even maybe disagree with ourselves, right, five minutes later. This is something that we academics are really, really good at, at saying, I have no clue, and yet I have a very strong opinion about something.

And the other thing that I think we need to bring into the conversation, again, to make sure that we get the handle on this responsible AI thing that is all of the rage, is we have to laugh about it. We have to laugh about ourselves. We have to bring humor into the conversation.

Because, you know, one of the salient differences between people and machines is that machines don't have a sense of humor, last I checked, and we do, right. So to make a topic human, let's be human in this.

So what is my definition of responsible AI? I don't really have a good definition, but I am going to try to try to at the same time agree and disagree with Will. So I'll agree in that I have also been

thinking about what this Center for Responsible AI, and what it is, and what our mission is, is to make ourselves obsolete in a few years, to make it so that responsible AI becomes synonymous with just AI and no one talks about responsible AI anymore.

So there we are in agreement, but I disagree that responsible AI is just legally compliant AI. I think that there is much more to that conversation meaning that perhaps a better way maybe to put this is, how do we make AI socially sustainable? And this includes, of course, legal compliance, but it also includes us as people wearing whatever hats that we wear at the moment, an individual, a mother, an academic, a member of the public-at-large.

How do we make it so that we are able to participate and to exercise our ethics and to bring our values into the conversation about AI?

So to speak maybe a little bit about the role that I see for academics and for academic institutions in this space, of course, the primary mission of an academic institution like New York University is an educational mission and also a research mission. And in this topic of responsible AI, education and research coalesce and they reenforce each other.

Because when we start thinking about responsible AI and we start thinking about responsible technology, responsible computing, more generally, we are forced to make strides, have to make strides to both educate the practitioners, but also educate members of the public; and then there is also a lot of research that is at these interdisciplinary groups, disciplinary boundaries of technology, and policy, and law, and social psychology, and social science, et cetera, et cetera, so we are never bored.

Now I'm technologist, right. I am a computer scientist by training and I am an engineer. So, as engineers, we build, and I very much have an engineering mindset. And the style of research that I personally find most appealing is when we do research that is driven by practical humans. You know, nothing wrong with basic research, but I think that here we actually have a responsibility to respond to what the world is asking of us right now.

And, in my view, the biggest gap right now is not in the availability where unavailability of sophisticated data collection and data analysis managed to be very extreme here to make an extreme point, might even say we don't need new algorithms right now. And this must sound really strange coming from a computer scientist where developing new algorithms is our bread and butter. This is how

we publish and therefore sur survive, right.

So we don't need new algorithms. But what we do need is to understand how to make the algorithms that we have equity aware. And this is a term that I actually prefer to responsible these days. How can we invent ethics and policy constraints, for example, into the systems that we have

And, here again, I will agree with Will, who said, well, legally compliant, right, we have laws that exist but maybe we need to change laws. So I'm also contradicting myself when I say that we don't need new algorithms because, of course, just taking an algorithm such as this and fixing it a little bit to now become legally compliant is not going to work; instead, really we need to rethink the entire stack.

And then there is another gap, of course, that is very important and that we have to fill with our educational mission. And that is, how do we make people algorithm aware? How do we teach people at different levels about what algorithms can and cannot do?

And how do we create this deliberation with as many stakeholders as possible at the table about what algorithms should and shouldn't do? What shouldn't we be asking algorithms to do?

So this is responsible AI at New York University, at NYU Tandon School of Engineering, specifically focuses on the applied research, where the goal is to build these systems that embed equity.

As a primary objective, the work we do is never purely that technical and it always involves education of data scientists, of policymakers, or members of the public. And it involves the research and the practice of public participation and public engagement and a deep commitment to helping shape policy which we have so far been doing mostly at the local level in New York City. And the goal there is really to help make technology work for all of us, not only for the select few.

MS. TURNER LEE: yes, yes. So, yeah, I mean, look, I want to go -- you know, Will, I want to give you this opportunity respond back and then I have a couple of questions I think that will trail into Will and Natasha. Because I think you bring up a lot of really interesting things, Julia, which in my work is actually similar which is having these become much more participatory models that we actually develop and the extent to which we actually look at improving upon the performance of algorithms versus creating ones that continue to skew.

But I do want to go back to Will on this whole legal compliance thing because I do think that there is a difference between building algorithms from the engineering side that you hope are going

to be legally compliance, or building algorithms -- so, for example, some of the pragmatic and practical examples is building algorithms that exclude.

Because when you build it to exclude from a technical cadence, right, you then go out of sort with advertising, or fair credit housing, employment loss. So, Will, I just want to give you a chance to respond. And then, Natasha, I want to come to you about this whole thing.

Because part of this conversation we're having for the three of you is about fairness. Co I want to come back to you and really talk about, can we really define fairness, which actually leads to the big goal, the big question mark?

Will?

MR. HURD: Well, look, I though Julia's point on making humans algorithm aware is important.

MS. TURNER LEE: Yeah.

MR. HURD: Because we're going to have AI that's going to be able to write codes, okay. So once you have that then, you know, what should you be using it for? And these are some basic questions. It's funny the longer I go in my career in public policy, the more basic my questions become.

And so, it's like, what is the role of the United States and the rest of the world, right? What percentage of GDP should be used on, you know, healthcare? Like these are some simple questions, but to get to an answer is really hard. And that's why we need engineers taking classical liberal arts, you know, education.

We need people that are in liberal arts taking some, you know, some basic software development classes and understanding some of these concepts because the two things are interrelated and you have to have people that have a grasp of both of those in order to answer some of those questions.

And, look I don't disagree with anything that Julia said. I think, you know, the laws that we have is a starting point and if we need to make things better, you know, that's the standard by which we evaluating things.

And so I actually believe that artificial intelligence could be helping us make better decision in applying the law in a more fair way rather than being the use case that manipulates that. That

would be a golden, you know, that would be the -- if I had my magic wand, on a position that we actually get to.

MS. TURNER LEE: Yeah. And that's why I love this conversation. I always tell people at Brookings and the external world that the longest paper that I ever wrote was with an engineer, a sociologist, myself, and a lawyer. (Laughter) That paper felt like it never ended, right. But it was actually the best paper that I wrote because I actually got to see things from their perspective.

And, Natasha, one of those things that we actually -- and another point, too, is that we are argued about was this whole idea of fairness because fairness is so elusive. When you think about fairness as a term, I mean, how do you operationalize that?

Particularly from an industry perspective, what do you do to make fairness look like a metric versus, you know, something that people want to strive for? Then I'll have Julia and Will jump in.

MS. CRAMPTON: Thanks, Nicol, I really do think fairness is a perfect principle to use to illustrate these things that we have just been talking about. It is, of course, one of our AI principles. In fact, it's the first of them. It's a core focus of our responsible AI standard. It's part of our culture. And, yes, from my perspective, it represents one of the hardest and most pressing challenges in this space, particularly when it comes to try and to operationalize the concept.

So we have teams working through different types of fairness hubs, here at Microsoft. We focus on three types of hubs: quality of service hubs, so this would be -- an example of this is where you might have a facial recognition system that works really well for some demographic groups and poorly for others; allocation hub, so you can imagine this as something that is sort of unfairly allocates employment opportunities to one group at a higher rate than others; and then we also try and think about representational hub, so this might be when a model generates output that reenforces stereotypes, or it continues to sort of over or underrepresent certain groups.

And we really start with having teams think about deeply, in a way that they may not have in the past, who their stakeholders are for the particular system in that building. And the way that we do that is through impact assessment, and we ask teams to think broadly about where, and how, and for whom the system is being.

So they can build out a picture of, you know, what things might go right, but probably

more importantly what things could go wrong and for whom. You know, we didn't have teams having identified some demographic groups that might be affected by the system.

We do have them think very carefully about the datasets that they are using to train and teach their algorithms. And we've got, you know, various tools that we use to try and analyze the composition of those datasets because we want to make sure that, you know, that the datasets reflect the diversity of the world in which we live.

But it goes beyond data, you know, sometimes the discussion about fairness really heavily cinches just on data. But, as Julia was alluding to, when we're building models we make lots of choices about features, about model structures, about data functions, and about training algorithms. And in each of those decisions, design decisions that we take, we need to prioritize fairness at every point.

We do have some tooling that we use to try and help making, detecting, and mitigating fairness issues earlier, so we have got a couple of open-source choices that we have developed here at Microsoft. One is called (inaudible); and the other is called Era Analysis.

And they do help with certain fairness challenges, but I think it's fair to say that the tooling in this area is still pretty nascent and there is a lot of work as a community that we need to do to try and build that out further.

Now, of course, testing needs to be done in the real world, in the context of which the system is going to operate and not just in the labs because it's very often the case that results in the lab don't reflect the reality of the range of scenarios that really occur in the wild. And you have got to keep testing throughout deployment to make sure that you're uncovering issues that might emerge in the real world.

And I think critically, you know, decisions that we take when we are trying to build out fairness across this life cycle do require transparency because people at the end of those systems need to understand them. And to Julia's point, the limitations that they place on the system when you take student decisions as you're building out a system.

So this is sort of overview of a life cycle-based approach that we try and take here at Microsoft. But in the spirit of acknowledging what Julia said about the challenges yet, and what we know, and what we don't know, there is certainly a few areas that I would point to that are very real in present

challenges for operationalizing fairness.

The first thing I would say is that fairness is deeply contextual, as has been discussed in much of the literature in this space. It is not a single definition of fairness that works well in all circumstances and that context is really critical.

It matters how and when the system what we used. So while, of course, a general additive (phonetic) of a program such as the one that I run at Microsoft is just sort of deeply study particular scenarios and see whether you can generalize out and find a passion that you can apply to other scenarios.

Fairness doesn't always lend itself to that type of generalization. And it was found that, you know, things that we have learned in the context of working on the fairness of our facial recognition systems doesn't always generalize out even to, say, face-to-cheek system; and so that contextual nature of fairness is really something that is critical to recognize but it's also a challenge in that space.

I would also say that there are very few norms around methodologies and measurement right now when it comes to fairness, and I think actually the responsible AI space more generally. So when our engineers are working on fairness testing they ae eager to use industry standard methodologies or to understand what good or excellent looks like, in terms of the fair performance systems; that there aren't really standard metrics for that today.

A lot of the AI benchmarks that we see are quite focused on technical performance aspects. So, you know, they might refer to accuracy in the sense of precision and recall, but we really need to together build out these benchmarks to make sure that they better take into account the sort of sociotechnical considerations that we're trying to measure. How can we weave the fair performance, the transparent performance of a model benchmarks that we hold up as industry standards?

And then, finally, I would just say that there are certain types of fairness issues that are still genuinely research problems. So I mentioned earlier that we work on quality of service harms. Those are quite well understood in terms of how we might be able to identify and mitigate those sorts of harms.

But representational harms, these are the situations where a model might generate stereotypes or over or underrepresent groups. There are still lots of open research questions there. And

we really shouldn't even try, in my opinion, to move to mitigating those sorts of harms until we deeply understand what the challenges are.

So I think I'd close by saying that, you know, we need to continue to recognize where research questions lie, and we need to support the research community in the efforts to help us better understand them.

MS. TURNER LEE: Yeah, I mean, I think -- Natasha, thank you for laying that out, this whole idea that fairness is contextualized because I actually say the same thing, that there are not enough norms.

And this kind of goes back I think to what Julie was saying that it kind of intersects with the reputational side. Some of the norms are mainstream norms. They're not even built on the lived experiences of the people that are effected by the technology.

And then, I would say, third, yes, and it kind of goes back to Will said, being as a legislator, he's looking at the black box outcome. And I think wat Julie was sort of suggesting, you have got to start with equity. So I think what everyone has said so far has been interesting.

And, Julie, I want to come to you, right. Because one of the areas, in addition to facial recognition that we see a lot of this discussion is in employment AI. And that is becoming, I think, the big pressure cooker when we think about.

And I love it, people, that you all are using the word, "socio," because my mother always wondered what I was going to do with a sociology degree. (Laughter) But the fact that we can look at employment algorithms and see that these proxies of face, and zip code are being used.

As a sociologist, we studied that years ago. We identified Black-sounding names, like, Latanya Sweeney's work suggests, or zip code, or the fact that now you can see a person's face doesn't provide for blind interview.

So talk to me a little bit about that because I know, Julie, in New York, you're actually working on that and globally with your work, in terms of putting fairness into employment AI and hiring.

MS. STOYANOVICH: Yes, happy to, but I also wanted to ask Will, did you have something quick to add before we dive into this new area?

MR. HURD: Thank you, Julia. The only thing I was going to say is what complicates

everything that Natasha said is you're not going to get any direction for government on some of these questions, right. So all of the companies are having to develop these things themselves which means you have to be putting it what is best for the customer.

And it's not always what's best for the companies. And I'll use, you know, opt-in versus opt-out. Yes, as a company, you want someone has to opt-out because you got them, right. But nobody who has ever clicked unsubscribe thought that their information was going to stay there.

And then, you know, unsubscribe me and you can hit me in nine more months, right. And so those are some of the decisions that businesses that are developing these are going to have to make. But thank you, Julia.

MS. STOYANOVICH: Absolutely. And this time maybe we have trouble disagreeing with anything that either Natasha or Will said -- (laughter) -- so let's see if we can stir up a bit more controversy going forward.

So I want to pick up from where Natasha left off with that allocation cards. And this, of course, is a very nice way to categorize the different types of harms, right, Natasha, that you already discussed. And as much as we can do to categorize things to try and put things into boxes, you know, we should always attempt to do that while still of course recognizing that these boundaries within boxes are not very strict.

So when talking about fairness in allocation cards, I think that we have been overly focusing on fairness in outcomes, who gets hired and who doesn't get hired. Do we have enough women or enough people of color being represented among the finalists, for example, in a job interview round?

And we have been assuming that in an effort to fix -- and I'm going to use air quotes here, "bias in outcomes," it is somehow justified to make the process by which we arrive at these outcomes arbitrary and uncontestable.

At a slightly ore general level, we have been operationalizing fairness as something that is a property of a dataset or of an algorithmic process. And this is not at all a productive point of view. And this is something that Natasha spoke about as well already by discussing that in life cycle view of what made the decision systems and fairness as it pertains to the entire life cycle and as it interacts with transparency and explainability, right.

But another life cycle that is important here, in addition to the data collection data analysis deployment is also this life cycle, the design, development, deployment, and oversight of automated decision systems.

So when we think about fairness, I do really believe that it's not productive for us to be limiting ourselves to a frog's eye view of dataset, model, output. And then you'll notice that something happens where your output is gender biased or it's racist, as in not selecting enough people of a particular demographic or socioeconomic group; and then the only thing you can do, if that's your world view, is tweak the output, tweak the box, tweak the input, right, and this really limits what you can do.

So another way to put this is that this myopic sort of view it's not productive because it dismisses the before and the after, the socio, legal, technical context, in which these systems operate.

So let me give an example now, of course, prompted by Nicol, and thank you for this prompt, and that is the use of algorithmic hiring tools. These tools, the promise that they make, is that hiring and job search -- so hiring for employers and job search for job seekers is going to be made more efficient, less paperwork.

You somehow are going to be able to find candidates who are well-qualified for your job very quickly and you, as a job seeker, are going to be able to find all positions that are a good fit with a click of a button.

But, importantly, we also hear this argument that the use of algorithmic tools will help us improve workforce diversity. And this is, the argument goes, is because humans are biased when they hire and there is plenty of evidence to that effect. We have no choice but to use machines to step in and hire on our behalf. And this, in my opinion, is a very, very dangerous premise. It's a very dangerous point of view to hold.

And, normally, this is where I go into my soliloquy that today I will avoid, about how -- because humans are biased, and then when we hired we're going to embed essentially the bias that they're hiring decisions have been having on the world, in the data that we get; and then because our data exhibits this bias, we trade in algorithms and they will give us biased models, et cetera, et cetera, right.

But, ultimately, of course, it's not very easy to do bias algorithms. It's not very easy to do

bias datasets because we just don't understand the dataset doesn't know whether it's biased and for what reason, Is it a bad reflection of a good world, a good reflection of a bad world, or are these distortions compounding, right?

But so, the point here though is that although when we talk and think about bias a lot, even if a tool is designed that helps us improve the diversity of the applicants being considered, let's say, at some step in the hiring process if, for example, it admits a sufficient number of candidates of each required demographic or socioeconomic group, but the decisions of this tool are otherwise arbitrary then can this tool really be considered fair?

What if these decisions are worse than arbitrary? What if the tool is picking up signal like a person's disability status and we have no way to know that this is happening?

What if it's measuring something about a job applicant that we have no reason than to be believe to be relevant for the job for which they are applying, like, are they able to pop balloons quickly enough in a video game, right? I mean why is that relevant?

And so where do we go from here? I think where we go is that an integral part of this conversation about fairness is also making sure that the tools that we are building and deploying actually work, by some definition of work. And these tools, of course, are engineering artifacts, right.

We build them to some specification. And so we should not take the gleam that they work on face. Nobody who has ever built a car or a bridge would say it works, believe me, you have to actually show that it doesn't crash and it doesn't endanger the lives of people, right.

So to know whether algorithmic tools work we should similarly use the scientific method and that is very simple. We formulate the hypothesis that states in a falsifiable way that the tool indeed, in this case, selects employees who do well on the job and who in fact do well on the job in a way that is better than if you were selecting them with a random coin flip, or a bunch of random coin flicks.

So comparing performance to performance of essentially a random coin flip is the lowest bar. And this is not something for which we have right now a standard. And I think that that's just really worrisome.

And then, beyond the definition of works, which is just better than a coin flip, we should also, of course, pay attention to the impacts that these tools have on different stakeholders. And Natasha

spoke about this already with impact assessments, right.

Fairness of outcomes by gender or by race, which is what our legal frameworks will tell us that we should be compliant with, is not sufficient. It's not a sufficiently high bar. What about intersectional groups? What about Black women, right?

We don't currently have a requirement that tells us to acquire enough Black women or enough disabled people of color. What about individuals with disabilities, right? They may not disclose and they often don't disclose their disability status when they apply for a job. And so we cannot audit for bias in this case because we just don't have that information.

MS. TURNER LEE: Right, that's right.

MS. STOYANOVICH: Yep. And so we cannot guess algorithm, what the actual or potential harms of such tools are by just having just some people at the table. No matter how smart these people are and how dedicated they are, we have to incentivize and solicit input much more broadly.

MS. TURNER LEE: Natasha, I want to stay on this, right. Because I think what Julia did was she just brought us back to square one, in terms of how difficult and complicated this space is. And then, Will, I have a really super question for you that I want to ask you, as a legislator.

But, Natasha, how would you respond to that? At the end of the day, no matter how you put it, it seems like fairness is going to continue to be elusive. But I know you have given us a framework for at least how we can measure what that fairness looks like. I mean are we talking more about a lean towards standards or, you know, any response to what Julia said?

MS. CRAMPTON: I do think we need to move towards standards. Ad I think Julia's idea about as sort of putting forth hypotheses that can be falsified and tested against is a good one, you know.

And now that similar approach that we try and take internally is that, you know, we try and ask teams to make sure that they have evidence to support the claimed benefits of their AI systems.

And, you know, sometimes this captures people by surprise because they sort of think, well, but of course my AI system with its, you know, magical powers is going to be better than a human being. But by actually asking for the rigor of the evidence that's a really good way of testing with a particular system it's fit for purpose.

So I do think this reenforces the need for standards. And the only way that we are going

to be able to figure out how to, you know, develop those standards is to really roll up our sleeves and I think work through these hard problems together in a participatory way that is really essential here.

So I think we are going to have to find the sort of exemplar use cases that illustrate different types of challenges and really roll up our sleeves to figure out how we might be able to make inroads here.

Now we are unlikely to be able to get to a state of fiction (phonetic), but is nonetheless a very important task. And I think by getting to standards will be able to start to build public confidence in the systems, and we should be able to provide that kind of assurance.

You know, we should be able to have third party auditors to be able to assist, you know, the processes that companies such as Microsoft adopts to make sure that we have got a reverse approach to these issues and to be able to essentially prove that we are walking the talk and not just talking the talk.

And that's, you know, absolutely something that we do for privacy or security. And so we have really just got to do this hard, immediate work right now of really trying to peel back the layers of the onion they are trying to find what some of these standards might look like.

MS. TURNER LEE: Yeah, just really a quick reminder, if you have questions, events@brookings.edu, or please provide it on #AIBias. You know, this conversation, I could stay here all day. In fact, I want to take my seat and just become a participant because I think this has become a really interesting conversation unpacking some of the challenges.

But, Will, I want to go to you. Because, as a legislator, you said something earlier which is, you know, you can really hook on to this idea of consumer algorithmic literacy, right. But when you start to breakdown the models that have been presented where does the legislator fall in this? How do they begin to unpack? What part of it do they actually legislate?

I heard Julia say something about outcomes, right, in terms of inputs versus outcomes. So I just want to hear where you're standing because I know that was your challenge with you often.

MR. HURD: For sure. So I think it's, like, one of the places that I would start with is you have to be able to show how the algorithm made the decision, like, I know it's hard. Everybody says, oh, these things are so complicated. But if that algorithm is going to have an impact on people, you're going

to have to be able to explain why this happened.

And because that company is going to get dragged in front of Congress and you're going to have some engineer be, like, uh, we don't know, right, that this is just not going to work. Now when it comes to in research developments, I think that's a different scenario.

But when it comes to actually deploying to consumers, then you have to be able to explain why that decision was made, right; and then is the point in which there is potential regulatory involvements or legislation, is it whenever that Schuman made the decision, right, in that process. And it could be as far back in the design of the system or where and I think that is one area.

But this is so complicated, like, we -- look, in Congress, we couldn't even get agreement on when somebody should be notified when there was a breach. It took Robin and I, Robin Kelley and I, three years to pass a piece of legislation on the internet of things when everybody agreed on what the outcome was, right.

So that's, you know, trying to get even further down into the development of the methods or the tools. I think that's pretty hard. And guess what? I don't think we want Congress to mess up those, you know, to get in the way and slow down the development of these tools.

And my final point is when we do talk about the fairness of outcomes, you know, in some of these to understand bias you have got to collect data that in some places you can't collect, right. In the EU, you can't collect on an application whether you are a Black female in some cases.

So if you're not able to collect some of this data to ensure that there is equity in the process and how these tools, how are you going to get to that end goal of -- what was the phrase that was used, the equity, being equity aware, if you haven't collected some of that data on the frontend.

And I don't know the answer to that. I'd love someone's insights on that, and I bet you Julia has an opinion.

MS. TURNER LEE: Yeah. And, Julia, if you can, just step in, too. And, Natasha, I'll throw this to you and then we'll go to Q&A from the audience on the EU. I mean I think what Will is also describing is the EU scenario where they have become much more prescriptive, not just in the proxy standard, but they have also been prescriptive around AI.

So I'm just curious, is that where we're leaning in this conversation, or are we going to

keep some fluidity for the type of self-regulatory moves that we can take on this issue, as well as some participatory involvement from the people who are being effected? So, Julia, I'll go to you, and then Natasha.

MS. STOYANOVICH: Yeah, so just very quickly. There are limits to any single method of instilling responsible AI and responsible technology requisites, right. So auditing certainly is a useful tool in the toolkit, but auditing is not going to take us all of the way.

Because as we already all mentioned, in the U.S., as well, people cannot be compelled to disclose their membership in protected groups, right -- so disability status, age, gender, race, all of these things is not something that we can ask -- we can ask, but we cannot compel people to disclose.

So the way that I like to think about this is in terms of creating distributed accountability structures, strengthening our collective accountability in making sure, to help make sure that the way that we use technology, AI, in particular, is responsible.

So returning to the definition of responsible AI, whose responsibility is it? It's every single one of us, but it's not computer risk, right. Computers actually cannot take responsibility. They don't have free will, so they cannot be held accountable for things going right or wrong.

But we each have to participate in keeping these systems in check. And I will give just very quickly an example of how this might work that also takes us and kind of connects the thread with the standards conversation. I do think that we need standards. But more than standards for fairness, I think we need standards for disclosure, for public disclosure, specifically, about the use of algorithmic systems that impact people.

And a way that I like to think about this is using this metaphor of an additional label where we show to an individual being impacted, for example, by an algorithmic hiring system; that they weren't hired because they weren't able to tell apart quickly enough red squares from green squares on the screen; and then the person can speak up and say two things.

First of all, I'm color blind and so you are discriminating against me based on my disability. And, secondly, this is not something that is relevant for the job. My job does not require me to tell apart green shapes from red shapes, and so this gives a person an actual way to contest the decision, both in terms of relevance, but also in terms of discrimination.

So I think that when we work towards standards, the standards we should be thinking about is standards for public disclosure, standards for public education, developing a standardized process for agreeing in a society about what we should be exposing about it all. What definition of fairness matters to us? So that's my two cents.

MS. TURNER LEE: Yeah. Natasha, jump in, and I want you to pick up the EU-U.S. questions because I think that's been actually bubbling among our listeners as well. So they have standards when it comes to AI. I think many of us at Brookings are still getting through the report.

Can you discuss the EU regulations on the AI and how that's going to impact the U.S. as well?

MS. CRAMPTON: Sure. One of the observations I had, as I was working my way through the European pretzel is that there are actually some parallels between the approach in Europe and some of the approaches that we have seen bubble up here in the U.S.

And, like those approaches, I mean, are either the algorithmic accountability that Senators Wyden and Booker, and Representative Clark advanced, and we have seen sort of similar counterparts arise in some of the states as well.

You know, both of those types of proposals, plus the European one, really center on high-risk systems and there is actually a degree of some transatlantic overlap in the types of systems that are considered as high-risk.

And there also seems to be a bit of implicit behavior shown by the size of the Atlantic that regulation should fill gaps in existing little frameworks. Now the size and shape of those gaps is pretty different in the U.S. and in Europe. But I don't think this idea of a gap analysis and building from existing law is the right approach. I think the European proposal thing goes deeper in a couple of respects.

So, first of all, it does start to tackle this notion that to your advance of distribution is accountability because it does try to go into the allocation of responsibilities across different actors in the supply chain. And it certainly sets out much more specific sort of design time documentation and testing obligations that providers of high-risk systems have to follow.

You know, as we have been building our own internal practice of responsible AI, we have had to think about what those specific obligations look like. And when we have adopted those practices,

we have challenged ourselves to always keep asking a few questions.

One of those is, how will this practice lead to a more trustworthy system? Are these practices practical and proportionate and achievable? Do they actually generalize out the across different types of systems? And do they take account of likely future technological directions?

So I actually think this set of questions that we ask ourselves internally is probably quite an instructive set of questions for policymakers to ask us while including when we are looking at quite specific obligations because we want to be sure that those specific obligations are actually meeting their policy objectives.

Look, I think it probably too early to decide whether we will see another example of the Brussels effect, yeah. You know, on the one hand, European proposal clearly is an ambitious one that has some of the hallmarks of the GDPI effort.

But it has been really encouraging to see a real focus on transatlantic cooperation on AI policy. And I think that's something that wasn't really present at the time GDPI was being enacted.

You know, I do think we should take advantage of this current moment to ask sort of -- make sure that the impetus towards transatlantic cooperation and policy is honest. I think we should try and, you know, move to a world where we do have harmonized rules based on shared values.

And we don't really want to end up in a situation where we have, you know, conformity assessments or other forms of assurance that sort of have to be repeated across geographies. And I think there is a real opportunity here to try and create systems of mutual recognition and cooperation.

MS. TURNER LEE: Yeah. And I'm looking at one of the twitter questions from my friend, Mike Nelson, who says, "We also probably need to agree on what is AI because that's still something that is quite debated in the long-term."

And, Will, I'm going to turn this question over to you. How can philanthropy, professional, and trade associations play a constructive and meaningful role in advancing a responsible AI agenda? Because I know, you know, you have thought about that, as well as other stakeholders coming to the table.

MR. HURD: Sure. And let me start this answer with two broader points that we haven't addressed because I think they need to be said. We won't be able to define the rules if we don't own the

technology, right.

And so we are in a race and the broader question is, can the United States and our allies keep up with the greatest geopolitical adversary which is, in my opinion, the Chinese government, without becoming (inaudible) in surveillance tape, right. Like, this is the broad issue and we have to get these answers right so that, you know, responsible and ethical AI is what drives this into the future.

And what other outside groups, philanthropic groups, trade organizations, is start explaining and showing to elected officials how artificial intelligence is impacting, in a positive way, or a negative way, your own industries. And because I think every industry is different, how this unfolds in banking is different from construction.

And so I think, you know, educating those members, educating those staff, ,you know, anybody who is involved in this, have you gone to your local member of Congress and talked to the District Director or the Chief of Staff of that individual to start improving the basic education?

It seems like such a simple thing to say, but there is dearth of understanding up in Washington, D.C., and that's how these groups can help educate our elected officials on these issues.

MS. TURNER LEE: Thank you for that. Julia, I'm going to give you this question. This question is: As long as humans are involved is ethical AI even a possibility? And if you can answer that in about 60 seconds because we're running out of time. We could have added, like, easily another half hour or hour to this conversation.

MS. STOYANOVICH: The answer is yes. And, yes, humans are always involved and ethics is something arises through deliberation, right. I mean it's ethics that humans deliberate on. It's our values. It's our responsibility. So, absolutely, without human involvement I don't think we can have ethical anything including ethical AI, yep.

MS. TURNER LEE: Yeah. Anybody else want to just respond to that, in terms of, without humans is ethical even a possibility, as we get ready to wrap up?

MR. HURD: I'm trying to achieve Julia's, you know, original request of being -- of some conflict, but I agree with her 100%.

(Laughter)

MS. TURNER LEE: Yeah. I do that as a moderator. I think I have pushed you down to

the deepest level of inquiry and then we go from there. And I appreciate all of the listeners, too, with questions.

I will ask one other question, and again, and then I'll wrap up. Natasha, trustworthy versus responsible AI, sort of like that human to the ethics, trustworthy and responsible, is there a reason why we need to use responsible more than trustworthy?

MS. CRAMPTON: Look, we have chosen to use responsible at Microsoft because we think it really sort of communicates clearly that it requires action on our part. You know, even at the most senior levels of the company, President Brad Smith, you know, if you're creating technology that changes the world, then you have a responsibility to help address the world that you have helped to create.

So I think, you know, responsibility really does convey that there are acts, there are things that you need to do, and that there are consequences for not doing those things and carrying out what your obligations are.

So I do think ultimately, you know, there is that they are closely intertwined. I mean we also believe clearly that, you know, people don't use technology that they don't trust. But I think the notion of responsible does resonate for us because it extends to cover concepts as well, like, reliability and safety which some people might not think of being, you know, traditionally kind of ethical constructs though there is of course a debate there as well. So for us it really encapsulates the suite of issues plus our own obligation to take action.

MS. TURNER LEE: Yeah. Well, I want to say thank you to all three of you. And before you all log off, let's just say, thank you in a virtual applause to the three of you for taking your time.

And I want to say to everybody who was watching, Brookings is talking about these issues at the Center for Technology, which I run. We actually have an AI stream where you can find on the Brookings website a series of papers that deal with everything from governance to bias and national security.

So I would ask you to actually look at that. We have just published a new paper today. We also have a podcast, the TechTank podcast, where we actually talk about these issues as well, and continue these conversations as we also look at how that intersects with what policymakers are thinking.

But, most importantly, in my own work, I think I am very, very humbled to have the three

of these folks here today. Because I think at the end of the day responsible AI is just one framework for looking at it. And, as you have heard, underneath that are a range of adjectives. And the only one I would add that we didn't hear too much about is inclusivity.

And as we go forward, let's make sure that that inclusiveness is actually part of the framing. And I'll just put a shameless plug, my energy star rating chapter is coming out soon so that you can actually see what I'm talking about with that.

So, with that, thank you for enjoying the afternoon with us. Thank you to everybody that is here. And we will see you at the next podcast. Will, Julia, and Natasha, thank you.

* * * * *

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, 2024