

THE BROOKINGS INSTITUTION

WEBINAR

HOW INCENTIVE-BASED RATING SYSTEMS
CAN MITIGATE ONLINE BIASES

Washington, D.C.

Monday, October 17, 2022

PARTICIPANTS:

Moderator:

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

Panelists:

JOHN BREYAULT
Vice President of Public Policy, Telecommunications, and Fraud, National Consumers League

RUSSELL C. WALD
Director of Policy, Stanford Institute for Human-Centered Artificial Intelligence

MIRIAM VOGEL
President and CEO, EqualAI

* * * * *

ANDERSON COURT REPORTING
1800 Diagonal Road, Suite 600
Alexandria, VA 22314
Phone (703) 519-7180 Fax (703) 519-7190

P R O C E E D I N G S

MS. TURNER LEE: Good morning, everyone. How are you? I'm Dr. Nicole Turner Lee. I'm a senior fellow in Governance Studies and the director of the Center for Technology Innovation at the Brookings Institution. Well, you are in the right place this morning because we're going to talk about a really important topic, which is artificial intelligence, particularly algorithmic bias and some mitigation strategies.

Many of you who have been following my work have seen that I'm really focused on trying to figure out how do we identify and mitigate online biases? What do we do to really discern whether or not those risks are consumer preferences or they're preferences that we need to be concerned about? In the U.K. they call it high risk and low risk algorithms in financial services, employment, education, healthcare, among other categories.

We know here in the U.S. we're just making a lot of inroads to make sure that we come up with fair and responsible AI. But one of the things that I've been sort of struggling with is what are some pragmatic strategies that we can actually embody that take on a sociotechnical approach? And so, as part of this Webinar this morning, I'm so delighted to be joined by some friends who are going to discuss to a certain extent a recent book chapter I've put out in an AI and governance handbook for the Oxford University Press.

And this particular chapter really piqued my curiosity on what would be some ways to give consumers more agency over these algorithmic models that may have the potential to create great consequence, whether implicit consequence or explicitly. And so, today we're going to talk a little bit about what that model looks like. It's very much shaped around something you've heard me say in the past, this Energy Star rating approach to algorithmic fairness. It involves a collaborative participatory effort among industry, government agencies, consumer groups, and other associations who basically want to come up with some governance standards on what makes fair and responsible AI.

But going a little further, much like the Energy Star rating program that applies to consumer appliances, it's actually led by the Federal Trade Commission, the Environmental Protection

Agency, among other agencies. What I'd like to see is us deploy some type of signature, whether it's a Better Housekeeping Seal approach, whether it's some type of framing or standards that people understand that when this algorithm is optimized for its particular function, it's not going to discriminate nor bias against particular users.

And on top of that, what I really enjoyed about doing the research on this paper and talking about the Energy Star rating, it allows for consumer agency. How many of us really get to critique whether or not those algorithms apply to us? We don't. And so, we're in an internet culture where we can actually apply some incentive-based ratings, letting people who develop these frameworks understand that technically this may work in the lab, but when it's actually deployed in the context of my experience, it may create disparate impact or differential treatment.

So, I started this exercise. And this chapter's available again through an Oxford University handbook on AI and governance. Please look at the chapter among the other titles that are actually part of this conversation. And we at Brookings, we're going to continue this year to push the envelope on pragmatic solutions. Look out for my new framework. It's called an anti-racist approach to AI. And I'm excited about the next level of this work, which is really putting tools in the toolbox so we can make progress.

So, today, like I said I'm joined by some friends. And I'm really excited about it because they all have been sort of thinking whether in the middle or on the edge of this idea of giving people back control over their algorithms. We're joined by my friend, John Breyault, who is the vice president of Public Policy, Telecommunications, and Fraud at the National Consumers League. Hey, John.

MR. BREYAULT: Hey, Nicol.

MS. TURNER LEE: Miriam Vogel, who's the president and CEO of EqualAI. Thank you, my friend, for coming on.

MS. VOGEL: Thank you, Nicol.

MS. TURNER LEE: And Russell Wald, who is the director of policy at Stanford University's Institute for Human-Centered Artificial Intelligence. And, Russell, I was watching you on

Twitter for that fabulous hearing that you hosted for the U.S. Government. Great to see you.

So, with that, John, I'm going to start with you because, obviously my model assumes some level of consumer protection, right, leads with consumers. You know, we talk about humans, I'm really trying to center consumers at the heart of this model. You work on their behalf. And for the most part, I'm thinking about as you heard me say this Better Housekeeping Seal, you know, Jeanie Barton from back in the day --

MR. BREYAULT: Yeah.

MS. TURNER LEE: -- that was something that she used to always talk about, where's the Better Housekeeping Seal when it comes to algorithmic fairness. Talk to us a little bit about your thoughts on how AI emboldens consumers in its use. So, where do they fit and what can be done to ensure that they have increased agency in the design, development, and deployment of these models? So, I'll start with you.

MR. BREYAULT: Sure. Well, you know, consumers tend to affect the marketplace through their pocketbooks. It's what they're buying, and what they're using, and how companies can make money off of them. And so, to the extent that consumers have agency over AI models, it really comes down to which models are they using.

But AI is a really interesting use case because we often don't know as consumers that, you know, AI is even being used to affect how we look at the marketplace. So, for example, the ads that we might see when we're online, are definitely impacted by AI. The fact that we're not getting spam in our email boxes is definitely impacted by AI that filters out bad emails or spam emails.

But how consumers can actually impact that I think is a really interesting question. I tend to be a little bit skeptical when we talk about sort of, you know, consumers exercising agency over things that they don't really understand to begin with. So, I know all of us are familiar with, you know, our privacy preferences and things that we get on Facebook and other places. And as a consumer myself, it often gets a little bit frustrating that sort of we're asked -- we're being asked increasingly to turn that dial or flip that switch to somehow affect our experience and we're not really sure if it actually does.

So, you know, I'm not really sure that consumers are right now in a place where they have the skills or the information they need to actually apply their agency to those models. So, I think there's a lot more learning that needs to happen. I think they need to be -- we need more information out there about, you know, how is this model actually built? What is it doing to affect me on a day-to-day basis? And especially, to put all that information in a way that consumers can actually understand because, you know, we're not all computer scientists. We're not all building AI models ourselves. So, actually turning that into plain English is a real challenge I think for the development community.

MS. TURNER LEE: You know, I love that because it is a hard question. In fact, when I was writing this article, I was thinking am I really getting at what I want to say, right, in this chapter because to your point, this is like a virtual cycle to a certain extent. There's on and off switch. And the more that we go deeper in our digital profiles and composite behaviors, we basically don't know, you know, what variables are being scraped to actually make these determinations about us.

But I want to get back to that because the question becomes do we sort of throw our hands up and say that consumers should not have any part in this, right? And at some point, we got to figure out particularly -- and John, you and I talk about this all the time -- when people are subjected to algorithmic models that create differential treatment or worse, desperate impact, how do they actually respond to that? Like what recourse do they have?

And so, I think once we're talking through this, I really want to make sure we give our listening audience like how then do we engage people in these feedback loops so that developers even do a better job? So, I love that. So, put a pin in it because we're going to come back to that because that was one of the hardest parts in really figuring out where consumers should come in.

Before I get to my friend, Miriam, I want to make sure that you all are following us on Twitter. Please use the #OnlineBiasRatings, OnlineBiasRatings. And we're going to have time for Q&A. You could throw those questions up in Twitter or you can send it to events@brookings.edu. So, we want to hear your questions because we'll have time to actually listen to all of you as well.

So, Miriam, let's talk about what you've done at your organization, which is the reputation

badge. So, I think this is a close way of trying to operationalize something that allows for more participatory discussion and dialogue around these conversations of transparency and fairness. Tell us a little bit more about the EqualAI Badge Program and how it actually drives some level of accountability and even tell us how you got there, right, so we kind of understand the origins.

MS. VOGEL: It would be my pleasure. But I have to start out by thanking Nicol for hosting this conversation. I'm so excited that you're really digging in on the how-tos so we can take action because we don't have time to wait. So, thank you for hosting this conversation. And likewise, I need to reiterate my thanks to Russell for hosting NIAC last week. Really helpful starts that you both are fostering to make sure we get to answers and quickly.

So, thank you for that question. The Badge Program is something I have been so excited about. It was not something EqualAI originally planned to create. What happens is like you we go around talking to companies trying to create engagement. My message is you are now an AI company. You are now an AI organization because you are using AI in pivotal ways. And so, while you think you create products and services, you actually have the responsibilities of an AI company at this point.

Well, then the question is, well, so what do I do? Because as we both know, there's no clear national and international standards today. And I'm sympathetic. I'm a former general counsel. So, my orientation is, okay, how do we figure out the best practices and the models because as you know, we can't wait a few years for the frameworks to all be clear. We can't wait for the regulatory path to be settled. AI is being created and deployed at scale not just today, but it has been for years, and it'll be so deeply intertwined. We need to make sure that people have guidance that they can rely on today.

And so, we reached out to several experts in AI governance and AI frameworks. And we together built this series of six talks. For senior executives at companies what we realized is you have to get senior executive buy-in first. So, we targeted that audience because if you went to others in the company where you didn't have the C-suite buy-in, it would become frustrating because they really need the senior level buy-in. As you know, you need to have accountability at the highest levels and buy-in. It takes some time, resources, and commitment.

So, we bring this group of participants who really care about responsible AI governance and we across six sessions exposed them to Cathy O'Neil, the brilliant author of "Weapons of Mass Destruction", who in so many ways opened this field, who helps create an ethical AI matrix. Here's the concrete ways that you dig in and try and find bias in your AI systems and other risks. I mean, that's the interesting part. We're looking for bias. What kind of bias? Well, you look for all of them in the same types of formats with the same times of frameworks.

Likewise, you're looking for cybersecurity threats. You're looking for risks that will impede your system. We talked to -- we bring in other experts from Google and Microsoft and Salesforce who share their best practices for responsible AI governance. And something that I know is clear to us but is less clear to others and I really love it, one of my favorite parts, is bringing in a panel with lawyers and a panel of policymakers. The lawyers saying first of all, here's how we're advising our clients. Here are the risks and liabilities we're seeing right now. We're also trying to create the idea of a lawyer as a partner in this effort because they need to be. And there's a lot of work we're doing there I can talk about another time.

And the final piece is with policymakers. They are shaping the horizon. They're establishing the guardrails. And so, the more we can understand what's happening in real time, the better they can make their planning. So, it's over 6-1/2 months. We added a new session so there can be more participant-led engagement. Like you said, it has to be participatory. We've been asked to make our sessions bigger. You know, they're about 20, 25. We haven't been able to yet because they are so participant involved.

We have communities that we're building. They're learning to rely on one another. They're learning how to implement these tools under Chatham House Rules so that they can talk about what challenges they are all facing. And at the end, we've built this community and people who can rely on each other as well as the experts we've introduced them to. And very timely. We're about to launch our next cohort October 27th.

MS. TURNER LEE: Nice. So, do they leave with a certificate of completion? Or is there

something that, you know, because this was part of like the thinking of this application of the Energy Star rating, this visual representation that people have accomplished this, right? Or have thought through these questions.

MS. VOGEL: Great question. Yes, they get a virtual certificate. They also get online badges that they can promote on LinkedIn and others to demonstrate that they've made this commitment over six months.

MS. TURNER LEE: Love it. Love it. Russell, jumping into you, right, this conversation is interesting because you're more of a technologist than all of us on this call, right? I mean to a certain extent.

MR. WALD: Not quite. I just work with a lot.

MS. TURNER LEE: Not quite, but close, right? You're at Stanford, so by association we'll put you there. But when you think about what we're talking about so far and just the problems of really enacting the type of agency. And I love the way this conversation is going. It's agency of consumer. It's agency of industry. You talk about human agency. Is this what you're referring to when we start to mitigate some of these risks that we're discussing in terms of the human concept? Or is there more to it to ensure much more transparent and responsible AI?

MR. WALD: Yeah, so, when we're looking at human agency, I suppose we're looking I think one part is starting this early, and getting in early, and stating those values, and understanding those values pretty early. So, if you look at the lifecycle of deployed automated systems, it's a patchwork of layers that are intended right now to probably stop bias or do something along those lines and keep a clear focus on what is human-centered.

And for us, human-centered has righted concepts. But one is, it's developed to the benefit of humans. Two, it is intended to augment human capability, not replace it. And three, it's the intelligent design is what's found in nature and trying to copy humans so that it can work more clearly with humans in that sense.

And let me give you a good example of one thing that we've done that we hope is kind of

starting to create another one of those layers in that patchwork. And one of those is we have now, within Stanford, we're a grant making organization with to faculty internally. And so, they have to apply for one of our grants. And if they had human subjects, you would have to do an institutional review board, right, as anybody does in this case.

Well, irrespective of whether you have human subjects or not, you also with us have to do what's called an ethics and society review statement. It is that early layer of asking yourself the questions of who is a part of your dataset. What are you doing in these particular cases? Who is on your team helping develop this? What are the downstream effects of all of this technology and where that goes?

And the important part of that is is at least it's one early layer that you've put in to get at that point of that early part of the development phase, and they will be another series of layers. So, it's not perfect. But again, it's if you put in your Energy Star rating, you put in an ethics review -- ethics and society review statement at the beginning and you have these other layers, you can hopefully slow down these biases that are starting to creep in.

MS. TURNER LEE: Yeah, and that's the thing, I mean, I know that with my technologist friends, they're often like, well, this slows down innovation if you start putting in a human review board, you start thinking about, you know, the governance standards. And one thing for people to know, and I'm going to go to John because I know he's going to have a different opinion on this.

MR. WALD: Could I just say something real quick --

MS. TURNER LEE: Yeah, go ahead. Go ahead, Russell.

MR. WALD: -- Nicol, on that? I just want to say a lot of our faculty, you know, people say how it slows down innovation --

MS. TURNER LEE: Yeah, yeah.

MR. WALD: -- well, our faculty welcomed it, ironically.

MS. TURNER LEE: Right.

MR. WALD: We thought that we would be irritating them with another layer of like an

IRB, and it turns out that some of them have actually felt a bit of relief to be stimulated and think that far in advance of what the consequences are of the technology. Sorry, I just wanted to say that --

MS. TURNER LEE: No, no, and I --

MR. WALD: -- it's not -- I think there are people who --

MS. TURNER LEE: -- I think that's great.

MR. WALD: -- welcome it in some cases.

MS. TURNER LEE: That's right. No, I actually am glad that you said that because I think part of the pushback sometimes comes with the innovation community itself. But what you're suggesting, and this is what I think all of us on this Webinar have suggested, you're really trying to also mitigate reputational risk. So, the more that you can as an industry through the reputation badge, stay off the front page of the newspaper because you put out pretty much "bad AI", it's worth the challenge. And I actually propose it's worth it in high-risk categories like financial services, employment, education, and healthcare. Where the consequences are very hard once they are forfeited, right, to go back and sort of regress back to where you were.

Reinstate that. I mean, there's so many countless examples and we are not using this Webinar today to go through the examples. People need to go get my chapter, read the chapter. Because we all know all the panelists say the same examples. But at the same time, we need solutions just to build better systems.

So, John, I want to come back to you because a lot of this is self-regulatory. And you and I go back and forth. Is this something that should be, again, going back to your comment around consumer agency, you know, where they fit in this process, how they operationalize that. It's also a self-regulatory framework I'm proposing, which you and I know has its good and its bad, right. So, tell me a little bit about does that actually impact this ability of these models to effective in the long run?

MR. BREYALD: Yeah, so, you know, Nicol, I think you're right that there are good self-regulatory regimes and there are bad self-regulatory regimes. And unfortunately, what I often see is self-regulatory regimes that are created by companies or organizations that have vested interests in avoiding

more stringent regulation. And they put up these self-regulatory regimes, promote them as doing so great, but when actually dig into how the policies within them are created, it's often a very insular process that tends to help them game -- the entrenched interests that create the system help them game the system to avoid sort of costly changes to their business models.

And so, you know, in that sense, you know, I appreciate that you're putting forward a model like Energy Star because it is widely recognized. It's different, I think, than other self-regulatory models because it does involve the participation and the buy-in of government agencies who ultimately have the ability to regulate if, you know, a self-regulatory model starts to be abused.

But, you know, I want to know that Energy Star is, you know, while it's certainly one I think we all understand, I can certainly click a button on homedepot.com to only washers and dryers that have the Energy Star rating. You know, the end result of that process, that certification process is a product that a consumer can choose to purchase or not. I can get that refrigerator. I can get that washing machine. I can get that dryer. Or I don't have to.

The problem when you're talking about AI is that consumers often have very little agency over whether and how an AI model affects them. And I think you were alluding to this a little bit in your previous statement. Things like criminal justice when AI models are applied to sentencing, for example. A consumer really doesn't have much agency for whether they can choose that AI sentence or they prefer to go for the non-AI sentence.

So, you know, I think what I might encourage a look at as an alternative perhaps to this Energy Star model is something closer to what Underwriters Laboratory does. And if folks aren't familiar with UL, just go find any electrical appliance in your house and look at the tag on the back and you'll probably see a UL certification. What UL does is that they consider the safety of a particular product. Usually, it's around things like fire hazards or other safety that goes into building materials or electronics.

And once that certification is there, the consumer doesn't really have to look for it, right? It's assumed it's there. That standard is so widely accepted that government building codes, for example, and insurance companies often require it. So, the consumer doesn't have to exercise agency over

whether their coffeemaker is safe from causing a fire in their house or not. It's just assumed that it will be.

And so, I think I apply that UL model to what you were talking about earlier, which is sort of this high-risk versus low-risk sort of AI, right? And I certainly think the case for a model where we just assume that the AI does not have detrimental biases is going to be baked into high-risk AI. Like that seems to me to be the right path to go. So, the consumers don't even have to think about that.

Low risk, maybe there's more of a gray area, maybe there's more of an opportunity for sort of these self-regulatory regimes like Energy Star to have more of an impact. But certainly, I think, you know, we need to make sure number one, that if consumers are going to be expected to exercise agency over something, it needs to be something that they recognize that that's even necessary for them to do.

MS. TURNER LEE: Yeah, no, I love that. I want to see if anybody wants to jump in on what John said before I ask another question. Anybody have feedback on that? Miriam?

MS. VOGEL: Yeah, I think we can get more into it I assume we will. But I do think at the end of the day, I'll just, you know, not bury the lead, I think there has to be a balance. I think what the benefit of the Energy Star or the other models that John references it's an agreed upon model that brings clarity to those making a decision. And it's never going to be perfect.

You know, with nutrition labels I was on a project at the White House where we were trying to update it. It is never going to be perfect. Let me put it this way. But it's a standard statement. I love your idea because it just brings clarity to this very complicated discussion. And when you're thinking about government versus internal regulation, self-regulation, I think it's a false choice. I think it has to be both.

MS. TURNER LEE: Mm-hmm, mm-hmm. Russell?

MR. WALD: Yeah, if I could add to some of this. Whether it is a Energy Star based style system or a UL, either/or, I think they're both very good ideas. I think the one concern I have about this, and I hope we spend some time talking about it, and I don't want to deviate too far from the conversation, but a need for public sector AI.

MS. TURNER LEE: Yes.

MR. WALD: And having a more increased ability of talent coming into the federal government in this particular case. The concern here is even if you are -- even if this program we wanted to say let's do it tomorrow, we couldn't. We don't have the talent available to be able to do this type of investigative work whether it is on the Energy Star side of certifying these companies or being able to -- we couldn't scale it right now if we wanted to.

MS. TURNER LEE: Right.

MR. WALD: And a big part of that is we just don't have a trained workforce ready for this. And so, what we really need to do is I think invest in that next generation now and start getting people -- giving people tools and access and the ability to learn how to do this type of work. Create a set of norms, professional norms around it, and then go from there. Without that workforce, we won't be -- these are great ideas, but they won't get very far.

MS. TURNER LEE: Right. So, I'm going to pushback on that just a little bit. Let me ask chime in a little a bit.

MR. WALD: Good. Now, it's going to get even more fun.

MS. TURNER LEE: Like you're talking about a workforce too in that article. So, two things that I want to sort of first stay with the whole concept of having some type of pragmatic solution. I see it like this. So, a couple weeks ago, the White House came out with the AI Bill of Rights, which is really the standard governance of what our federal government wants to see integrated into our public agencies. And many of us had the opportunity to really read through that document and see how it's going to be endorsed by a variety of federal agencies who care about fair AI. So, I really want to really applaud people like Alondra Nelson, Dr. Nelson, and others who worked very diligently on that model.

The key thing though is, and this goes back to John's point, the extent to which a bill of rights applies to people, or does it apply to company, right? And so, the extent to which those are best practices that companies or government entities should actually be deployed. So, we got that one bucket.

Then we have NIST, who does a variety of useability studies on various technologies that come out of their shop that have to be industry approved like facial recognition, which I think does a great job of providing some type of technical standard. The challenge is, and I like the way John is sort of pushing it a little bit on the UL standard, is that we really, it's sort of a combination of the two that allows people to know one, that there's consumer confidence that -- let's go with the UL standard -- that when I plug in this computer, it is not going to burn down my house, basically. Because it actually has the specifications necessary to create the type of AI that may be in high-risk scenarios.

So, Russell, I'm going to your point, then I'm going to open it up. On the Energy Star rating, the way I thought about it is it gives consumers the opportunity to know that there has been some standard setting, but it also brings consumers, industry, government agencies all into the mix to determine what things, what are blind spots, what are we missing. We have a lot of government procurement of AI, and we find ourselves, all of us on this call, when we look at those models, basically government sort of retracting and saying, well, I didn't know this. I didn't know that. I should have thought about this. Because now they're going and they're fixing up whatever consequence has been placed on their actual constituents.

So, what we're trying to get at or what I've been trying to get at is a visible model that basically marries the technical cadence with the operational expectation of these models when they actually go in the field. I think we're mature enough to know that if I apply for a loan online, that I should actually not be rejected because of my race, or the assumption of my race based on the AI. Companies don't want to land up on the front page of the newspaper.

But I like what you're talking about, Russell, because I think that the challenge is you don't have enough smart people who know the technical side, but you do have enough smart people who know the sociological side. And the way that we actually build a workforce that is representative of the lived experiences and disciplines that go into this may make it the case where we don't have to wait until we're ready to actually deploy these models, we may have the right people in place.

So, that's one of the reasons why you all saw in my article I don't suggest that we build a

technical workforce. I suggest that we build a sociotechnical workforce of sociologists, and philosophers, and lawyers as Miriam has done, because those are the people that pretty much have to interrogate before these models go to market whether or not there are going to be blind spots.

So, I'll shut up. But I was going to get to your question on workforce. Because I think that's one thing that has stopped us from actually moving forward as Miriam said with some of these things because we're waiting for the conditions to be right, and the technology's not always right. Go ahead, Miriam, and I'll go to Russell.

MS. VOGEL: Thank you. And please never shut up. Please keep talking. Thank you for that.

MS. TURNER LEE: Well, I'm going to say I'm the moderator, I'm acting like a panelist. So, I'm going to pull back.

MS. VOGEL: No, please. It would be a shame if you didn't. No, but I just also wanted to get back to the point of the reason I said whichever model you choose, you know, they're both going to work, let's go forward, is because, you know, I guess it depends on what you're trying to solve with this model. If you're trying to build assurance that there is no problems with this AI system, you know, we're going to have to spend a lot more time thinking about what this looks like. I don't know if we'll get to that answer.

As we know, AI will iterate, learn new patterns, create new potential harms. We're talking about the full variety of risks like I mentioned. You're talking about risks of bias based on race and gender, which we are thinking about. But what about the cyber? What about age? What about geography? On and on and on and on. So, there's going to be continual iterations to our benefit and our detriment.

And I think where I'm coming from is right now, we have a standard that -- well, we don't have a standard. We have a system that in some ways disincentives doing this important work of the voluntary --

MS. TURNER LEE: Yes.

MS. VOGEL: -- checking. You know, Russell, and you brought up the interesting conversation about whether or not it's an impediment to innovation. But at the end of the day from a legal standpoint, you're creating a document demonstrating where you have potential vulnerabilities and liabilities that can be used against you. So, what's the incentive there? What I like about just having standard documentation that people who are savvy enough to understand the code and the variables, great, they'll understand it.

And in layman's terms, it says here's what they checked. Here are the standards that people said we need to check for. Basics, you know, which populations are over and underserved? You know, you bring up in your book, the chapter the great point about looking where is there diversity? Where is there not? Trying to understand which populations we did not account for. What were the use cases it was built for? Where wasn't it?

So, that at the bottom line, end of the day, there is both some clarity as to what and was not envisioned. Simplistic for sure but at least a starting place. And information for those down the line as we get more specific. And an expectation changing the narrative so that it is an expectation for companies to participate not just because you're a good actor. Truly giving it the weight that it deserves to be doing this verification.

MS. TURNER LEE: Russell?

MR. WALD: Yeah, I just want to say I -- your point on the sociologists being a part of the entire group of people enforcing or reviewing are a part of this, is a really good concept and idea. And I think that multidisciplinary nature is what is desperately needed. So, it should not just be technologists in the room doing this review.

We do need to still train, I think, an overall workforce whether that be sociologists who in order to understand the technology well enough to do degrees of enforcement that would be required through the FTC I would presume like with the Energy Star system. So, I think I'm all for a multidisciplinary aspect here and we need more of that.

MS. TURNER LEE: Yeah, I mean, don't get me wrong. We do need more technologists.

I was talking to somebody the other day. I said should I go back to school to actually learn some more technology, right? Because I'm advocating for it. But maybe I need to learn it a little bit better. So, I think, you know, there are blind spots on both sides. But we just have to think about it an all-encompassing ecosystem, right? And what does that ecosystem for leadership.

MR. WALD: Yeah, I also think you were right too when you were saying that there is that side of, well, we can't do anything because we don't have the workforce in place. So, it does hamper like the willingness to really give it a shot or put things, good ideas out there. Almost as if you build it, they will come. Maybe if you build an Energy Star system, you could hire a lot of people who could do this, so.

MS. TURNER LEE: Yeah. Those of us on this call it's all about impact, right? So, John, jump in, please, on this conversation. But I want to keep it going because I think what we're getting at is maybe the operational side of this is really increasing consumer confidence, right, in these models. And just curious from you too, again, comment on this conversation, but also, what the role of disclosures might be.

MR. BREYALD: Well, you know, I tend to be somewhat skeptical of disclosures. And that's probably from far too many checkboxes that I have clicked through over years of being faced with them every time I open a new app or go to a new website. So, you know, I've unfortunately found that disclosures are too often used to hide information that they know a consumer's not going to read that could be very detrimental to them in the long run. So, in that sense, you know, sort of before you use this, here's a disclosure so you know what you're getting into, I'm skeptical of.

What I'm less skeptical of is disclosure that actually gives consumers information that they can use at the time they're going to make a choice that impacts their welfare. So, if I'm using an app, for example, that is going to use AI to make a decision about me, like tell me at the point that that's going to happen when I have a choice to do it. Like that's useful disclosure.

But sort of telling me ahead of time in dense legalese in 10 pages all the bad things that could happen if I use this product, and oh, by the way, if I want to use the product, I have to accept this, not very useful at all. And I think that's fairly well recognized within the technology community and the

privacy community that's probably where I'm most familiar with it.

But I want to sort of, you know, add another layer to this. Sort of how consumers can exercise their agency over this. And I think, you know, one, you can actually look at this also from a little bit of an indirect way that consumers could exercise agency. You know, I often think of sort of the government's purchasing power as being one way that consumers can -- that sort of consumer welfare standards can be sort of test-driven in the marketplace.

So, one area I could see the standards that we're talking about today be useful and this was prompted by your conversation about the White House's AI Bill of Rights, is if the government itself couldn't buy AI solutions for things like procurement, for example, unless those met a certain agreed upon set of standards, that is one way to start to get these out into the marketplace. Because I think what I've heard from Miriam and from Russell is true, you know, the pace of this advancement and innovation in this sector is far outstripping our ability to hire people, to operationalize any kind of standards, and is happening too quickly for us to wait on sort of a year's long standard setting process to happen.

But I do think the government itself can take steps right now to try and make sure that what they themselves, you know, practice what you preach, in a sense, when it comes to AI. To sort of get these concepts out there and get them more normalized.

MS. TURNER LEE: Yeah, you know, you had me thinking and for a moment I was like I love that because I think for the people who work in government who are listening to this, you know, part of it is what can government do to increase their best practices when it comes to all that we're speaking about today? And again, you know, whether you call it Energy Star rating, whether it's a spinoff of the UL rating, for me it's how are you exercising the right cadence internally to make sure that you're also paying attention to those blind spots?

So, someone said it earlier and I like Miriam's actually operationalizing of EqualAI Badge. What certifications are your team going through? Here at Brookings, we give, you know, talks to government agencies on implicit and explicit bias to actually train your developers on what that looks like. How do you actually have a body. Because you all remember, we hired a lot of like fairness specialists in

the last couple of years like I'm the chief fairness person and they're sort of trying, they're in a bubble by themselves trying to figure this out.

I think what we're probably seeing, and John that was such a great point, perhaps we need government to actually take this on and think through who sits as part of the multidisciplinary collaborative group that says, hey, we're about to purchase educational AI. Have we checked all the boxes? Is a person here who represents urban and rural communities, different learners? Who is sitting at the table to check for desperate impact? Boom, boom, boom.

And then to your point, then I want to turn it over to Miriam because this is where it gets really hard, right? You need to have the toolbox of audits and civil rights assessments and other things to make sure you're not just throwing the AI out there, but it's actually operationalized itself to equity.

Miriam, we didn't talk yet about equity, right? Because part of the reason I wrote this article really had less to do with the AI that's giving me recommendations for movies, it has to do with the AI that's making determinations on my quality of life. How do you put the equity back into these models in ways that we actually have something to work with when we're trying to operationalize something that's very hard to do?

MS. VOGEL: Great question. That is what it's all about here. I think we all here -- I can't speak for you, John. I imagine I'm speaking for you, Russell. That we are AI enthusiasts. We are very excited about all the ways that AI is changing our life from the movie recommendations and news playlists as you say, but also lifesaving telemedicine procedures and scientific discoveries.

So, our goal here with EqualAI, and I know yours too, is to identify ways to make it more broadly useful. To make sure that the AI we're using does create all those efficiencies and benefits that we are dreaming of it doing without hurting people along the way. And I think one way that we talk about at EqualAI is that you have to imagine that bias and other risks will embed -- you can say can, but will, embed at each step of the AI lifecycle through each human touchpoint.

You know, you talked about the bias training you're doing at Brookings. Well, that's because we need to level-set that this is not a new medium that -- you know, some people bring implicit

bias when they're hearing about AI. Some people think it was the nightmare of alien takeover. And some people think of it as the panacea where it is this neutral computerized scientific model that will do everything equitably. Well, we certainly know that is not true. It will do what we program it to do.

And so, knowing that bias and other risks can embed at each human touchpoint from the ideation phase, who has the privilege of creating an AI solution? I mean, that is a rare group that has that benefit, that opportunity. So, imagine, you know, if more people had that opportunity, we're going to have different kinds of AI.

So, from the get-go, we are limiting the vantage point in who is having access. And I guess, you know, to sum it up briefly, we really take to the point Kathy O'Neil's very important message that the best thing you can do at each cycle particularly testing at the end and as we iterate testing again and again, imagine for whom could this fail. And as you talk about in your chapter, there is no getting around diversity.

You know, we talked about here diversity and expertise. That is so important. You need to have the social scientists. You need to have ethicists, lawyers, my bias. We are in the habit of creating frameworks, reducing liabilities. This is our jam. You want to have policymakers. You want to have these different lenses because someone can only dream, imagine as broadly as their experience and creativity allows. And so, the more shared lived experiences you are bringing along the AI lifecycle, the better your AI is going to be.

So, I don't know if that answers your question in the more concrete terms. But the more concrete version of the broader topics I just talked about are making sure essentially that you have good AI hygiene.

MS. TURNER LEE: Yes.

MS. VOGEL: Making sure that you have accountability, a plan in place that your entire organization is aware of that your C-suite has signed off on. A point of contact. And clear expectations of what will be documented and when so that down the pipeline you can see what was and wasn't tested. Have clear definitions of what that meant at the time of testing and knowing for the future for whom could

this fail. Who did we not see that we need to see in future use cases and future populations that are using this AI system.

MS. TURNER LEE: No, I love that. And, John, it kind of goes back to this whole idea of disclosures. Because you and I agree disclosures for the sake of disclosures is a disclosure, right, for the most part. What we're talking about is, and I always give this example when we talk about facial recognition technology, if facial recognition technology requires certain criteria for a person like me with dark skin to be able to have a photo that allows me to, you know, travel, or get benefits, or whatever the case, or certify my benefits, then you need to tell me that. You need to tell me that as part of my use of this product these are the optimal conditions.

Or if I'm using Credit Karma, just for example, no affiliation, no endorsement of the company, who's sort of telling me what my credit score is, but really, it's based on the FICO analysis of what we've actually determined are the determinant factors of adverse and positive conditions for credit, tell me that FICO -- Credit Karma is based on algorithmic decisions.

I think what happens is, and this sort of picks up on this whole equity conversation, you're so right we don't know where things start and stop. Consumers definitely don't have any kind of disclosure to know what the fallibilities are of the technology. And we assume that it's going to work well on everybody because that's what the technical cadence says.

So, I want to take what Miriam said, sort of push that back to you on how do we then get consumers in this -- you know, I'm not going to let that go. And then we'll go to questions. So, if you all have questions, I'm going to have one more question for Russell and then we're going to go to some Q&A. Go ahead, John.

MR. BREYAUULT: Well, I mean, honestly, I think if you put an AI disclosure like the type you just described in front of 100 consumers, the mere fact that they saw that an AI was going to be used to make a decision about them for that product, would probably scare off at least half of them. Just because I think that people have been conditioned through media and other sources to be worried about AI decisions that are being made about them.

Now, rightly or wrongly, I think that that is the way things happen now. So, and I think it's because they don't understand for all the reasons I think Russell has talked about. You know, people, we don't have a trained workforce because people don't know how AI comes together and is created in the first place. So, strangeness is scary to a lot of people.

At the same time, I'm not all that confident yet that we're going to get to a place where consumers understand the models enough to know that the model that was used to create this AI for this product was different enough from the model that was used to create this AI for a similar product, that I'm going to choose one over the other.

MS. TURNER LEE: Yeah, yeah, I get it. I get that, right.

MR. BREYALULT: I'm not sure that's going to be something we should expect consumers to be equipped to make a meaningful decision about in the future. Now, your model may be the way to get to that. It may be able to say rather than know that, you know, the sort of the equity review board was used on this and how does it make a difference as opposed to this one is going to make a difference for consumer decision making. Maybe having a star on there that's associated with some sort of process that consumers have been sort of trained to know and trust, sort of the Good Housekeeping Seal that you talked about. Maybe that will have benefits down the line.

Until we get to that point, though, --

MS. TURNER LEE: Yes.

MR. BREYALULT: -- I think I'm going to stick by where I want to be, which is on the UL just you know when you plug it in it's going to be safe.

MS. TURNER LEE: Yes.

MR. BREYALULT: Right? And you don't really have to think about whether it's going to be there or not. It's just going to be there. At least for that baseline level of protection, particularly for protecting classes of information they're already protected under things like civil rights law. To know that the AI's not going to violate that and be safe for those things, is I think the baseline that consumers should be able to expect no matter if they're choosing a product or a service that AI is being used to power or

they're being affected by a model that they have no agency over that's powered by AI.

MS. TURNER LEE: Right. So, it's like a UL, Energy Star program, right. Something like that.

MR. BREYALD: Something like that.

MS. TURNER LEE: Well, that's what we do at Brookings. So, John, I will quote you as I go deeper in this model.

But listen, Russell, I want to get to questions, but I do have this question for you because you sort of started this conversation around, you know, people are accepting of some of the time it takes with innovation just simply because they want to get it right. When you think about this conversation, and I love the way John put it, even if you put the disclosure, you'll scare away half of the people. I mean, at the end of the day, and Miriam used the word algorithmic hygiene, which I love. I've used it myself. It's almost like algorithmic integrity, right? And so, I would love to hear particularly the work that you all are doing, human-incentive AI, at the end of the day, should we all as actors in this new burgeoning economy really look to algorithmic integrity?

It's almost like when you guys remember -- I'm going to date myself -- when we went from HTTP to HTTPS to sort of suggest to people that this is a safe environment in which to transact commerce and financial transactions? Russell, are we going to get there? Would this integrity actually exist in these models so that it does make a difference, you know, societal wise and generally for consumers?

MR. WALD: Potentially. I think it really starts with the developers first.

MS. TURNER LEE: Mm-hmm.

MR. WALD: We have to get there much earlier. We have to kind of create what my colleague and our AI ethicist, Rob Reich, refers to is essentially set up community norms.

MS. TURNER LEE: Mm-hmm.

MR. WALD: Establishing a set of professional norms in a way, right? So, lawyers have the American Bar Association, which require and guides them. A really good example of this is CRISPR,

right? So, when Jennifer Doudna was, you know, creating this, at one point I think she asked herself what if Hitler had this, right? And so, what she did is she put these extraordinarily massive community norms on. And in fact, someone in China did violate those norms. And what happened? They were shunned from the community. No one would accept any of their papers. They were able to receive no funding. It really created this just kind of strong area of where people said, okay, this is where we will go and not go.

And I think we're in this rapid race right now and hopefully it's not a race to the bottom of, you know, these text to image models that are being released just without much thought of what that could mean or something along those lines. What I'm hopeful for is what someone like Rob was calling for before, this strong set of norms where people know you just don't go there. And this is how we would follow this, and you build a whole set of standards within the community to be able to prevent this from happening. Or at least create, again, when I was mentioning earlier, those layers.

MS. TURNER LEE: Yes.

MR. WALD: Extra, extra layers until by the time it does get to the consumer, it is in a much better condition than it would have been otherwise without those layers.

I also want to say too we're talking a lot about consumers, but there are points in this of where people -- this technology may be deployed -- the automated systems may be deployed beyond an actual consumer, right?

MS. TURNER LEE: Yes.

MR. WALD: If I go on and I want to apply for a job and I go in and there's a hiring algorithm there, am I really a consumer in that case? Or am I someone who wants to apply for that job and feels that's the case. And even the consumer of a loan from a bank, if it's the bank with the lowest interest rate and you really desperately need that right now, yes, you're a consumer, but is it really choice if you are subject to certain -- if you're denied availability and you therefore have to go to the other bank, right? It's not so market driven in that sense.

MS. TURNER LEE: Yeah. Well, that's what I like about it. You know, my friend, Renee

Comings, Dr. Comings at UBA, always talks about data trauma.

MR. WALD: Mm-hmm.

MS. TURNER LEE: And one thing that has been interesting to me in this conversation, which is why I think it was important for me to push the envelope here, is it's about consumer agency, but when I say that I think listening to you all it sort of helps perfect a little bit of my thinking on this. It's about moving consumers from attributes, data attributes. It's a real active participant in this ecosystem.

Because to your point, we have hiring algorithms that consumers are basically going through like you said, the motion of the applicant, and they don't know that they're already prescreened out because the AI has essentially said that they're not showing eye contact, or they're lighting was off so they might look scary. I mean Black and Brown men, you know, have a tendency not to get past the prescreening algorithmic architecture because of the fact that, you know, we've been talking about, who's designing it? Who has the power over it?

So, I think it's really important for us as we have this conversation -- I'm going to go to this question -- that we continue to interrogate and throw out these questions, right? Because we all know that this is normative. So, we don't have to go through that conversation anymore, right? It's like saying, you know, voting rights is equally available. No, it's not. It's normative. We have a really bad voting rights system here in the United States. So, let's figure out how to solve it with basically by charge in this.

So, okay, let me go to some questions because we're running out of time. This is from a few folks who basically asked the same question. I'm going to read it as it's written. AI is not magic. It's just getting computers to do work which is currently done by humans. People apply bias to the work they do and always have. Today AI is already better than humans at most tasks to which it is applied, facial recognition, medical diagnosis, predictions on data, et cetera, but it's not perfect and probably never will be.

So, what is the threshold we'll require to use AI? And will it have to be perfect or how much better than the human alternative? So, I throw that question out there. Thank you to a few folks

that asked a similar question. Thank you to our Catalina in our office collapsing that. Anybody want to answer that one? What is the threshold we'll require or should require to use AI? And will it have to be perfect or how much better than the human alternative to be used? John, I'll go to you.

MR. BREYALD: Sure. Well, you know, I hesitate to agree that AI is already better than humans at most of the tasks to which it's been applied. I think that's a bold statement and I'm not sure it's supported by the data but let's take it at face value. And, you know, will it have to be perfect? No. Will it have to be better than the human alternative? Maybe. What does better mean in that context?

Is it better at doing the thing for which it's been assigned to do? Maybe. Is it better in the sense that it avoids mistakes that it hasn't been trained to spot? So, the facial recognition would be a great, you know, is the AI model perfect at detecting the faces that are put in front of it? Is it better than what most humans could do? Potentially.

But when it makes errors, is the AI accountable? Do we know who to go to? I can't ask the AI's manager if, you know, when it screws up. These are all things that I think qualities that having humans do tasks that AI is being applied to end up in a different result, particularly when you're dealing with how consumers interact with it.

Now, should there be a threshold? You know, sure of course there should be a threshold. There should be a threshold for any task that we give to anyone. I don't let my kid, you know, do half their homework and say, oh, that's good enough. But, you know, where should that threshold be? I mean, I think at a minimum it should be doing as well as what humans can do at that task that's it's been assigned and avoiding sort of the unintended consequences that a human might catch that the AI can't.

MS. TURNER LEE: Mm-hmm. Anybody else want to jump in on that question?

MS. VOGEL: You know, I just want to point out what's probably the obvious here is that, you know, the danger in thinking about what it should look like is that AI is being deployed at scale. So, we're not talking about a hypothetical of what we would like it to look like -- or we are talking about hypothetical. We're not talking about how we will finally allow AI into our ecosystem. It is out there. I think one important point to think about is how are we going to rein it in.

And there are certain variables to that that are important. Some ways we can't. Some ways it is out there. Some ways it is embedded deeply in systems, and societies, and cultures now that would be hard. I mean, are you going to take Zoom away from people? No. I mean, that's been our lifeline.

But the other piece -- so, it's important to think about absolutely what expectations we have, but also the reality of the fact that it is out there. And so, what do we want to put in place to make it safer and better. And as we're talking about that, I think one place, you know, we're talking about hiring. We talked about the variables of monitoring an interview who is and is not weeded out. But I think it's important to open the aperture a little bit farther here and realize we're talking about proving a negative in many of these cases.

So, we don't just want to look at who has been discriminated against affirmatively. We also have to in this AI space think about who has not gotten the opportunity to participate in the first place? Who was not selected by that recruiter because the algorithm weeded them out because of their zip code or another not only illegal, but improper variable that keeps the employer from seeing the best candidates, keeps the employees, or potential employees, from potential opportunities. So, I just wanted to add those two points, but happy to talk further about what the idealistic and appropriate standards should look like.

MS. TURNER LEE: Yes.

MR. WALD: Let me just in real quick, Nicol.

MS. TURNER LEE: Yes.

MR. WALD: I know you might have one more question. But I just want to build on a little bit of what John talked about. I think we need to be really careful here and get out of this concept of AI will singularly come in and take over one particular area, will do something better than humans, and therefore, it's just going to be completely relied upon through an AI agent that will do something.

I have a colleague who said this really well recently. And he's a radiologist and he says, you know, the question will AI replace radiologists is the wrong question. And he says instead the right

answer is radiologists who use AI will replace radiologists who don't, right? And that's how I think we need -- so, when we're getting these kinds of questions, I think we need to kind of we have a duty to kind of like clarify these particular areas. And I don't think that there will be any just AI agent out there doing things without humans in the loop in that sense.

MS. TURNER LEE: No, and that's actually a good point. And I think for the person who put the question out, the threshold, all we have to do is look at what's happened in the European Union, right? And we could see how they're defining some of those hypothetical thresholds and then say to ourselves do we want those same standards here in the U.S. That's another panel that I'll be doing. But that's an interesting way to look at it.

All right. I have another question in this last three minutes. This is from Jeremy Treadwell. Bias is persistent in humans, so instead of using AI on humans (decisions of quality of life), how do we reimagine AI as a tool for humans protecting user privacy or cybersecurity safeguards? And basically, in that same question, in the next 10 years, what do you all see as a relationship between AI and privacy?

Let's kind of go out on that. So, why don't we go round robin. What is the relationship between AI and privacy? And this is that thing that I think Jeremy's mentioning AI not used on humans where we're the attribute, versus AI used by humans as tools to protect privacy, human rights, cybersecurity, et cetera. John, you want to start with you?

MR. BREYALD: I feel like you've been going to me first his whole time, so.

MS. TURNER LEE: Okay. I can go -- I can make my AI go --

MR. BREYALD: No, no, no, I'm --

MS. TURNER LEE: -- physically.

MR. BREYALD: I can add -- I just want to put --

MS. TURNER LEE: My AI was programmed to look at you first.

MR. BREYALD: Oh. So, you know, I really like what Russell had to say about the radiologist. I mean, I think that AI as a tool to make the work that people are already doing better and

more efficient is how AI will be used. Now, will that efficiency gain come at a cost of consumer privacy because AI does depend on, tends to depend on acquiring very large datasets in order to do what it does.

You know, that is I think the question that will be, you know, the relationship between AI and privacy. How can we make AI work as efficiently as possible without requiring so much information to go into the AI to power it? Sort of AI efficiency is something that I think, you know, I'll be looking for in the years to come.

MS. TURNER LEE: Yeah. Russell, we'll go to you.

MR. WALD: Yeah. On the privacy front, to be candid, I'm not overly optimistic that it gets better. And I think that we are in desperate need for some federal guidance on this and it would be quite useful. My colleague, Jen King, who does a lot of work on the privacy aspect and, you know, areas like dark patterns or where there is just this algorithmic kind of nefarious intention, there are still people out there trying to chip away at the privacy area. And it's very concerning. And I don't think it gets better before it gets a little bit worse to be honest.

MS. TURNER LEE: Thank you.

MR. WALD: Sorry to end on a pessimistic note.

MS. TURNER LEE: Yeah, you know it always lands here. Just like our simple -- there's so much to do. Miriam, your comment on this last question.

MS. VOGEL: No, I think it's a great point that we need to be thinking about privacy. While we don't have the federal legislation in place today, I think we can't give up. And I think that it gets back to the whole picture, the whole big scope of what we're trying to do here, and this is build trust. Build trust in our systems, build trust in the processes, build trust within our organizations. You do that internally making sure people know that you're looking at your AI and other skills in other capacities and how they are representing your trust, your trust building, your deservedness of that trust.

So, looking at privacy. Looking at due process. A lot of the examples in your chapter require government due process. So, as John said, government can and should be the model. And what a great way that is to help demonstrate accountability, transparency, and other necessary elements. So,

bottom line is AI will be -- is currently and will continue to be a way to help supplement our functions as we continue to build the trust, which is why developing these ratings and other standards to build that trust are so important.

MS. TURNER LEE: Well, thank you everybody. I really appreciate this conversation. I appreciate you all joining me. It's always a pleasure to be among really smart people like all of you here. And for all of you who actually joined us to hear this webinar, we are still pushing through on this conversation. Please send me your comments via email, Twitter, LinkedIn. I really want to hear a little bit more about how this model rubs with you. Love the feedback in terms of looking at the UL model verses the Energy Star. But I think the bottom line is we need something. Because what we're doing actually requires a little bit more interrogation.

I'm Dr. Nicol Turner Lee, director of the Center for Technology Innovation. You can find the article, again, it's in the "AI Governance Handbook," published by the Oxford University Press. Some of the articles, including mine, are online. So, please visit that site. You can find the reference actually in the event description. Thank you very much, everybody. I appreciate all of you and let's keep talking. Thank you.

MR. BREYALD: Thanks, Nicol.

MS. VOGEL: 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

ANDERSON COURT REPORTING
1800 Diagonal Road, Suite 600
Alexandria, VA 22314
Phone (703) 519-7180 Fax (703) 519-7190