

THE BROOKINGS INSTITUTION
The *TechTank* Podcast

“Measuring Black women’s trust in AI”



Monday, August 4, 2025

Guests:

RAJ KORPAN
Nonresident Fellow,
Governance Studies,
The Brookings Institution

ZARAH GUILLEMET
Student,
Marlborough School

Host:

NICOL TURNER LEE
Co-host, The TechTank Podcast;
Director, Center for Technology Innovation;
Senior Fellow,
Governance Studies,
The Brookings Institution

[music]

CO-HOST NICOL TURNER LEE [00:00:00] You're listening to TechTank, a bi-weekly podcast from the Brookings Institution, exploring the most consequential technology issues of our time. From racial bias and algorithms to the future of work, TechTank takes big ideas and makes them accessible. Welcome to the TechTank podcast. I am co-host, Nicol Turner Lee, senior fellow in governance studies and the director of the Center for Technology Innovation at the Brookings Institution. Listen, artificial intelligence has already become integrated into much of our lives and more. And more industries are looking to adopt the technology. Interactions with models or even robots might soon become commonplace in sectors such as medicine, finance, employment, and government service. You all have heard me talk a lot about artificial intelligence with my co-host, Darrell West, in this podcast. And guess what? We'll talk about it again today. Right now, I have the AI Equity Lab. If you don't know much about it, go to the Brookings web page. Look up AI Equity lab, and you'll find out the work that we're doing. But much of it goes to, how are different communities interacting with AI? And most importantly, is it being done responsibly? With care for those who probably have the most to lose if the technology goes awry. And even if we develop the safest models and standards, there's another aspect that is often largely ignored and that is this thing called trust. How much do we trust interacting with AI technologies? And this is the case for people like myself who tend to be marginalized in technological spheres and that would be Black women. Today, I'm joined by Raj Corpin, who's actually Assistant Professor of Computer Science at Hunter College and a non-resident fellow at the Brookings Institution. He's no stranger to this podcast. He has been on before. And his student from Marlboro High School, Zarah Guillemet, who has been working with Raj on this important question, which is around the public opinion of Black women when it comes to AI. I'm really excited about this. This is a different take on this, and I think that be getting us a really robust research that, for those of you listening today, might be of interest to you as well. Raj, Zarah, thank you for joining me.

GUEST RAJ KORPAN [00:02:35] Thank you so much for having me.

CO-HOST NICOL TURNER LEE [00:02:36] Thank you for having us. Yeah, you know, I mean, this is such an interesting topic. And when this came across my desk, I thought to myself, let's talk about it, right? I mean outside of the fact that we have seen some ebbs and flows when it comes to Black women generally. Just in February, over 100,000 Black women lost their jobs and you know we're listed in some of the unemployment figures. And today I keep thinking about what role AI could actually play

in helping with finding new opportunities, or building new businesses, et cetera, so I hope we can get into that a little bit. But you both undertook some new research around this topic, but more importantly, Raj, you've been researching ethics and technologies for quite some time, whether it's in AI models, computational models in particular, or robotics. But trust seems to be, before we go into Black women, trust seems to be at the center of a lot of these conversations, Raj. Tell me a little about what your research has been doing around trust, and then what questions you had going into this particular research around Black women's public opinion of AI.

GUEST RAJ KORPAN [00:03:43] Thank you, Nicol, that's a great question. And you're absolutely right that trust really underlies our use of AI systems, the way we interact with them and who gets to use it and how they use it. And I think there's not enough attention paid to trust because trust is hard to measure, first of all. There's so many different ways to think about trust. Trust is this multi-dimensional uh construct that is impossible to actually look at right you can't go into someone's brain and actually measure this is how much trust they have in this ai system it is highly contextual right what is the context that they're using it in um and how does that historical context that individuals context impact them their use or non-use of ai systems And so in my lab, we really do think about trust in AI and trust in robotics, particularly, and thinking about one, how do we measure this trust accurately in a way that understands this kind of complexity of trust? But then what do you actually do with that trust, right? And how do you measure when that trust fails and recover from it? So for example, we have a project going on right now about intelligent robots in stressful situations. How do people interact and trust with those robots when they give you the wrong information, right? We know AI models hallucinate all the time and they give false information. And so how do people react to that, particularly if it's embodied in a robot? And then how do we respond to that kind of failure that happens? And so what we've seen is a lot of the research that has to do with trust in AI has really been focused on the, let's say general population, right? The majority population that might be using these technologies. And often what happens then is marginalized communities or historically underserved communities are left out, right. We're not focused in on these specific communities and their challenges and how these violations of trust might affect them. Instead, there's this kind of like broad brush that's being painted. And so I hope we can we can slowly chip away at that and that was part of what motivated this research.

CO-HOST NICOL TURNER LEE [00:06:10] Yeah, that's so interesting to me. And I like the way you sort of talk about it, which is what I try to do in the AI Equity Lab, which is, how do we begin to understand the contextual frames from which people define and use

these technologies? But most importantly, how do we do more specific and granular research that informs just how this could be applied to very similarly situated circumstances or populations? Now, Zahra, I want to get to you. I'm just so impressed. Tell me before we start talking about your role in this project, like Marlboro School, tell me more about you.

GUEST ZARAH GUILLEMET [00:06:46] Want to know more about you. Yes, so I'm a rising senior at Mulbrough School in LA. I do outside of my research program, which is the Leonetti O'Connell Honors Research and Science Program, which is a program that my school funds that allowed me to undertake this project with Dr. Corbin. Outside of that, I do robotics, I lead theater tech, I leader on neurodivergent affinity group, and I'm student ambassador. I'm also an older sister. People tell me that comes across when I talk, so. And actually next year I am undertaking another research project and that one is centered around marine biomimetics and mechanical engineering.

CO-HOST NICOL TURNER LEE [00:07:30] Yeah, I love it. I'm an older sister too, so we have an affinity. Well, thank you for sharing your background. And I want to jump into this project that you did do with Dr. Corbin. Tell me a little bit about how you defied trust going into this product as well and what questions you had.

GUEST ZARAH GUILLEMET [00:07:48] I was a freshman when AI started really jumping into mass media with the advent of ChatGPT. And I noticed one thing that I picked up on was that the portrayal of Chat GPT in common media, so the news, for example, was the biggest one, but then also social media platforms in general, just informal human interaction. It was a very kind of wary take on what AI could potentially be capable of. So one thing I noticed was kind of a prevalence of. Fear-mongering of maybe a little bit of conspiracy theorizing that I saw and I was kind of curious as to why. And where I went to find out maybe get a little bit more of an idea of what the pulse of everyday humans were thinking about AI was I went my family members. I'm African American so my family members are as well and one thing that I got a lot of was this technology isn't designed for us. This technology isn't meant to make us feel safe or help us at all. Why would we bother using something that isn't going to be tailored to our needs? And I think trust is reciprocal, obviously. So if you trust somebody, you have to trust that, or if you've trust somebody you have to believe that they are going to be doing what's in your best interest. And in return, typically we try to repay that person's trust in us. So we try to do what might be in their best interest, AI is not a person. So we can't do that with artificial intelligence. I think the inability to connect with any sort of

perceived humanity that people or that may exist in AI is one of the things that provides a barrier between everyday humans and artificial intelligence. And then there's also something to be said for trauma. So consider if you're talking to another person and you trust them and they do something to betray you, the next time you see them, you might be a little bit less inclined to give them the same level of trust that you did prior to their betrayal. That's kind of the situation that we see with a lot of marginalized groups and artificial intelligence and specifically like in general the scientific research fields because of things like medical racism, techno-racism, abuses of marginalized communities in the scientific field. Marginalized communities now are a little bit less likely to put their trust in scientific institutions or researchers, or scientific creations like artificial intelligence, because their experience with the field as a whole is one of kind of.

CO-HOST NICOL TURNER LEE [00:10:10] Trust. So that is such an interesting observation and I'm hoping that you'll continue with your PhD. So I'm just saying we need another doctor. But Raj I want to come back because I think what Zahra is sort of talking about is there needs to be sort of this deconstruction of proximity to technologies and in this particular case what I find to be so interesting in the way she sort laid out the question, trauma, right, and how that impacts certain populations. Particularly those that are very close to it. Tell me a little bit about why that was an important question to explore with this population of Black women when it came to their opinion of AI. And more importantly, were there any surprises that you found that you'd like to share with the audience?

GUEST RAJ KORPAN [00:10:56] It's a great question. My personal belief and this kind of guides my research and the research of my students is that we really need to be community centered in the way we're thinking about AI and how AI is used. And so it was really a matter of thinking about what are the challenges within this specific community and a lot of that came from Zarah and her experiences, but also looking at the literature to understand that each community faces these kind of intersectional needs that are going to be different. And we want to use that lens as we start to approach different research questions. And so that was where this kind of idea of trauma, intergenerational trauma, or trauma related to institutions really resonated because even though I'm not a Black, I'm Asian American and LGBTQ in the LGBTQ community. I totally understand that perspective of distrust in institutions and having been historically marginalized in the system and society that we live in, right? And so that really told me there's something here that we need to look into further about how does this kind of trauma and media perception impact the way that this particular community of Black women might view and interact with AI systems. So that was kind of

what drove this. Now you asked, what's the kind of surprising result or interesting result we got out of this? Well, one, I think our hypothesis was confirmed, right, that there is this correlation between people who've experienced or lived through harm from institutions, particularly law enforcement, have an increased mistrust of AI as well. Correlation is not causation, but it's this interesting kind of maybe pattern that we notice that people who've maybe experienced harm from those kind of institutions also maybe expect harm from AI, which is absolutely and now an institution as well.

CO-HOST NICOL TURNER LEE [00:13:15] And that's so interesting because, I mean, as you all know, I wrote a book on digital access and I interview a variety of people from farmers to, you know, two women sitting in front of a public housing development in Syracuse. And throughout the book is this conversation on the trustworthiness of just general technology infrastructure, right? In terms of whether or not you have access to a phone or a computer, et cetera. But then undergirding that was always this question, like, I don't trust the internet. I don't trust the Internet, right. So I think that's a trauma. That finds itself sort of generationally hitting into the AI space. But more importantly, what I find so interesting about what you're discussing, and Zahra, you sort of laid it out, that the distrust that people have, this particular population, in institutions also translates into the perception around the technology that in many respects is not necessarily replacing the institution, right? It's augmenting the institution. So in your research, Zahara, I'll come to you, Like. Did you find a particular age of an individual where there was more distrust, you know, because some would argue that younger generations tend to be much more open to the technology versus older generations with their regional differences. Like, you now, where did you see some of the divergence when it came to the people that you were sort of researching on this time?

GUEST ZARAH GUILLEMET [00:14:37] So we did see a little bit of divergence along age lines, but not in the way most people would think. So in general, I feel like there is kind of a stereotype or a general understanding that the older people get the less they trust artificial intelligence because the less familiar they are with it. Our 30 to 40 age group actually showed the highest levels of institutional mistrust toward artificial intelligence, even higher than our oldest age group was, which was 60 to 70. And so what that showed us was that levels of act like institutional personal trauma play a lot more into whether or not somebody trusts artificial intelligence than age does, because we did actually see a pretty linear trajectory in terms of as levels of pre existing technological and scientific trauma increase levels of mistrust or artificial intelligence increase regardless of age. And so interestingly enough, in that coming back to the fact that we had a peak in trust for women ages 30 to 40? Young women! Thus far,

according to a Pew Research Center study from 2022, young Black women are the demographic of Black women most likely right now to report high levels of medical racism. And so that's obviously not directly related to artificial intelligence, but it is a scientific institution that younger Black women are being taught to mistrust because they've had potentially destructive experiences within the medical industry. And so that was interesting. And obviously like Dr. Corbin said, correlation isn't causation, but it was still an interesting connection to be able to make that mistrust toward the medical field could potentially be spilling over into younger Black women's mistrust of all types of science.

CO-HOST NICOL TURNER LEE [00:16:15] So I want to pick up on that, Raj, right? Because I know for people listening, it could be suggested, well, is that a discrete finding or is it something that could be explained by other variables when it comes to just people's distrust in medical institutions and. You know, I mean, part of the challenge we have with AI is that most of us do not know that AI is actually being used, right, within a health clinic or in the exam room, even though it's showing up with the notes from the doctor or the potential prescription interaction, et cetera. Talk to me a little bit about like, you know, and I know this is preliminary research, so I don't want to go further than where you both have gone so far. And for people who are listening, I think this is a very interesting topic to myself personally. So I hope to continue this conversation and we hope to publish something more at Brookings. But Raj, talk to me about that. How much of that is variance in this correlation between where people find themselves in a medical institution and how they report AI being as equally distrustful than a doctor?

GUEST RAJ KORPAN [00:17:27] Yeah, I think that's a really interesting point because you're right, this is preliminary work. We're gonna obviously do more follow-up research, collect more data and do some more qualitative work as well and bringing people in to interview them to kind of get some more nuance about their perspectives. But what Zarah was just talking about about this kind of higher level of mistrust in this 30 to 40 age range, I think preliminarily this is an interesting finding and we'd wanna dig into it more, but kind of what my understanding of that age group is that that is kind of the millennial group that has gone through kind of growing up almost with the advent of the internet and now the emergence of AI, right? And so it is somewhat distinct than people who are under 30 and maybe over 40, how they might be experiencing AI and this particular group in the middle might see the most to lose from this system. They're also in this kind of period of their life where Maybe they're building families, maybe they're thinking about long-term financial stability, all of that. But is that unique to this

population of Black women or is this kind of like all 30 to 40 year olds might feel this way? I think we need to do more research. And then to your point, I think there is research that shows that mistrust in one institution or experience of harm from one institution. Does cause overall increased levels of mistrust across many different types of institutions, right? And we've seen that kind of in the trust in our civil and government institutions over the last, let's say decade at this point, right, where harm from one kind of results in this feeling of potential harm from any of these things because as an individual, you feel like you don't have the power to be represented and express yourself to these large institutions that are inaccessible to you. And so whether that means like an experience of medical racism in a healthcare setting now results in increased mistrust of AI systems, I don't know if we can directly connect those two, but there's certainly going to be this cumulative effect where more and more harms. Compounded in one place are certainly going to increase skepticism and fear in other places.

CO-HOST NICOL TURNER LEE [00:19:57] Well, and that's interesting too, because I sort of equated in my work on bias and mitigation, right, to a couple of things. So in the health care setting, we already know that we have increasingly severe rates of disparities when it comes to Black women in health care, right? Whether it's being underdiagnosed or not diagnosed at all, when we look at, you know, some of the more chronic and severe disease. Like breast cancer or respiratory disease, where there is sort of this mix between Black women not necessarily showing up in critical trials to give us a better assessment of our health disposition and what could potentially be cures versus not being able to ask the right questions when we are in settings where we should be trying to get better information on our health, or just the history of Henrietta Lacks, right? Who was a Black woman who literally had her DNA used for cancer discovery but yet never got credit for that but also died. A poor woman just based on the misrepresentation of the experiment and we can go way back to the Tuskegee experiment, et cetera. So what dawns on me is that there is probably a market for this type of information to better understand how to de-bias systems. But most importantly, create, and I like the way Zarah said it, more training data sets that represent the lived experiences of impacted populations, which is why I was so excited to have this conversation. Because I think we all know that there are these biases that exist that present obstacles and barriers to full optimization of AI technologies by an assortment of community populations. But there's also those that have consequential outcomes that can result in life or death and health being one of them. I mean, when you think about this body of work that you're building and this incredible significance that it has now, I mean my first question is, do we need to just see more of this, right? In terms of really

understanding how different groups interact with AI, you know, for the betterment of their particular contextual application or context?

GUEST ZARAH GUILLEMET [00:22:17] I think absolutely there needs to be a wider field of, or more expansion within the field of studies like these, and I think my idea behind creating this was that it could serve as kind of a blueprint or a preliminary run of what a specifically demographically targeted public opinion survey of a marginalized group could look like, because every marginalized group has their own historical context, has their own specific experiences that are. Related to why they may specifically mistrust an institution that differ from group to group. And because of that, we can't just take this one study about Black American women and use it to represent all marginalized groups across the world who feel disenfranchised by the scientific community. We have to make an effort to continue to build this data so that when we train AI models, they're getting diverse data that represents their entire user base. And AI user bases are very diverse, especially because, like Nicole said, a lot of times when AI is incorporated into certain aspects of everyday life, for example, search engines. People don't fully understand or haven't fully come to terms with the fact that they are using artificial intelligence because it's kind of running in the background. They're not actively engaging with it, it's just, oh, I searched something on Google and I got AI mode. So a lot people don't qualify that necessarily as using artificial intelligent because they and choose it, like they didn't click the button. Or actively download an app to make sure that they were using artificial intelligence, but they are interacting with it anyway. And so that kind of unknowing user base is still a user base. And companies like Google, for example, have global user bases. And so, that means the people that are using their services represent a slice of the global population, not just the people who might be included in a limited data set. The example that I like to use is if you're training an AI to identify fruit and all you feed it, like all you train it on is pictures of blueberries and like the occasional great. The day you show it a pineapple, it's going to be like, what is this? I have no idea how to interact with this. I've never seen it before. Must not be fruit. And that's what we see happening with a lot of marginalized groups.

GUEST RAJ KORPAN [00:24:21] Zarah, you make an excellent point and it really comes down to the data that goes into these models, either not being representative of all people's lived experiences. It's also about who gets to make these models and who's in the room when those decisions are made. And we know there's been plenty of studies now that show that in the tech industry, but particularly people who work on AI. It is not a diverse set of people and often many marginalize that people are not included in those conversations. Right? And so if, if a person with that identity is not in the room

when these decisions are made. There's a much higher likelihood that their needs are not going to be addressed or included.

CO-HOST NICOL TURNER LEE [00:25:11] But Raj, I wanna bring up a question for the two of you though that I want you to think about though. There is also this presumption that the AI has left the station and that it's available. So like Zarah said, like we could be doing something on a search query tool and AI could submit it, right? And it's making it easier for some people to sort of digest that information. I mean, I'm working on this piece right now in terms of Black women's dislocation out of the workforce and the fact that AI could potentially be providing some free resources for career redevelopment, retraining, and replacement into the labor market. I mean, do we draw lines on this public opinion of areas in which AI has left the station and there are some applicable use cases that we may want to pay attention to? Are you basically arguing it's sort of a fine line and we need to be careful in how we use it overall?

GUEST RAJ KORPAN [00:26:09] I think, yes, there are definitely positive use cases and those use cases need to be developed very carefully and with the community in partnership, right? It's not a top down, here's a solution for you in this community and this is what we think you need, right. It should truly be bottom up from people in that community saying, here are my needs, Here are my challenges, for example, loss of job right in this current labor market. Specifically how do we retrain or assist people to kind of move into new worlds. I think at the same time, there is the negative side of the train has left the station, which is that kind of assumption or acceptance that this is how it has to be, that the bias is baked in and there's nothing we can do about it at this point or that any mitigation efforts are just kind of like putting band-aids on the edges and not truly fixing what's the real underlying issues. And I think I wanna challenge that. I think we still have the ability to actually make meaningful change so that we're not just moving around on the margins. So for example, I wanted to bring up two points. One is about bias, right? I recently, I have a group of students working this summer on understanding both the explicit bias but also the implicit bias in large language models. Often you're not going to be self-identifying to a large language model that I'm a gay man, for example, or that I am a Black woman, right? You're not typing that into chat GPT before you start to put in your query. And yet there's research that shows that based on the language you use and how you actually write out your prompts, it can infer your identity. And that inference is also influencing the output it gives you. And so we're really interested in the case where maybe it's not an explicit indication of someone's identity, but are there ways that implicitly you're telling it who you are and it's changing the way that it's responding to you? And so far, results show that it is very true that that is

happening. And that the results are vastly different based on who it thinks you are, right? And this is just purely based on how you type out your prompts. The second point I wanted to make is about the changing ways that AI is being used, particularly about it being used over longer and longer periods of time, right. Of course, you probably know about people building relationships. There was the tragedy of someone who died after. They created this artificial relationship. But there's no research that is being done because these tools are moving so fast and no one is pausing to say, what is the actual impact this is happening on our society, on individuals, individuals who have things like psychosis? Is this contributing to their delusions? There are several cases now where that has happened. It has pushed them into psychosis. Or thinking about the impact on the loneliness epidemic, that people are now turning to these models for friendship and for social connection when it's not a human being, right? It can be the perfect friend for you. It's never gonna criticize you. It's not gonna push you. It's just gonna tell you exactly what you want to hear, right? And in human friendships, that doesn't happen, right, Because we are our we need both positive and negative feedback, right? And that's how we build connections with others, but also as we try to better ourselves as well. And so that's where I feel like, yes, there are certainly positive places we can use this technology, but we are not spending nearly enough time researching all of these challenges.

CO-HOST NICOL TURNER LEE [00:30:18] Yeah, I agree with you. And I really like the way that you two have nuanced this. There's AI, which is available. But clearly what I think I'm hearing from you both is that in particular use cases among certain populations, there needs to just be a heightened sense that how they interact with the product or service that is enabled by AI will be different. And if it's different, we also need to continue to interrogate. The extent to which it does not create the type of harmful consequences or the persistence of trauma-related memory that some populations have had, which to me is quite interesting. I mean, Zahra, you live in a space, and this is sort of my last question for the two of you, where you are experiencing AI. You're also, as a student, experiencing the political world in which we live. And, you know, unfortunately right now the appetite for doing the type of granular research that you're referencing that was important to you as a subject of the technology has become somewhat minimized. I mean, is there something to say, if you could, to a policymaker on the importance of this type of research to better enable, you know, public policies and programs and literacy that makes sense for certain groups? Like, What would be important? Is it that Black women just need to know that AI exists and that these are some of the opportunities and limitations? Is it the policy makers need to persist in funding projects like this or industry needs to pay better attention? You have the microphone, my friend, right? And what would be your suggestion for how

we actually improve upon AI as it's currently, you know, as you have currently stated its circumstance.

GUEST ZARAH GUILLEMET [00:32:10] I think there's so many parts, so many components that would have to go into truly improving, globally improving the experience of what it is to be an AI user for all types of people, not just Black women. Part of it, like you said, is the responsibility of the industry. AI developers have to make an effort to make sure that they are training their AI with diverse sets of data, with respectful sets of data, because one thing about currently existing large language models is in general, they're feed data sets are a little bit indiscriminate. For example, ChatGPT pulls from by and large the entire internet. And you and I both know there are some wild things on the entire Internet. And so because of that, Chat GPT will sometimes give, not necessarily, and especially in its earlier days, would sometimes give potentially destructive or harmful responses. And obviously OpenAI has made an effort to kind of curb that in the two years, wow, two years. Two years, one and a half years since ChatGPT was released to the public, but there's still more work to be done. And so part of a developer's job, I would say part of their responsibility kind of in this crusade to make AI more equitable would be to start paying very close attention to what they're training the artificial intelligence on because garbage in, garbage out, right? That applies to computer science as a whole, and especially to artificial intelligence if you feed an AI. For lack of a better term, trash data, it's going to give you, it's gonna generate trash responses. It's going to give you trash output. And so I think that's kind of the developer side of responsibility. And then I think for policy makers, everyday people, one thing that we should really be conscious of is educating ourselves on how artificial intelligence works, where it's being used, making sure that we understand, not necessarily the intentions because AI can't have intentions. It is just code, but the intentions of the industries and the institutions that are producing AI models and the considerations that they may or may not have when creating their product. So I think one thing that's kind of been a hallmark of our current age is a sort of passive consumerism, essentially, we'll see things, we will buy things, look at things, watch things, listen to things, without really thinking about whether or not we actually want to be using them or watching them or buying them or listening to them. Um, passive scrolling is like that. Um, for a lot of people actually cookies on websites are like that. They just click accept. They don't read the privacy policy, um, creating accounts. I could go on. I'm not going to, but I think being really intentional and understanding user rights are there for a reason. You can opt in or out of using certain things, especially if you feel that. They're not a good representation of the way that you want to be treated. Especially if you. As a consumer, your rights aren't being respected, or as a consumer the experience you're getting from the product isn't what you want to have. There's no

law saying you have to settle, you can just stop engaging with it. And especially for a more online society where user engagement is absolutely currency. That's how monetized ads work. Developers especially ought to be interested, and I think a lot of them are, in consumer data, in consumer experience, because the entire way they make money is consumer use. And if a large enough amount of consumers do say, do make an effort to make it known that they're not happy with the experience they're getting, or they think it could be made better, developers in general, and this is something that I think the tech industry should work on, being more to consumer data, consumer opinion. Public opinion and updating their datasets as such. A loud enough consumer base can absolutely make a difference. And I think that is something that people should understand. Like as a consumer, I'm not saying the customer is always right, but as a customer, you do have a voice.

CO-HOST NICOL TURNER LEE [00:36:08] Well, I appreciate that. I was just trying to think of when you get out of college, I want you to call me because I think you've got a future in the data analytics space as well as in the technology space. You know, Raj, I want to first commend you for finding what I call Zarah, an unhidden figure, a young person for once that actually is able to sit at the grown people's table and talk about these issues. And for our listeners. That's really important. We've done this a couple of times. We had some young people come on and talk about social media. And I think it's really important to have a young person come on and talk their experiences with AI. Meeraj, I'll have you give the final word. I mean, policy makers need to hear what this type of research is gleaming. And just from your take on this, how much more needs to be done to, as Zarah said, to ensure that we're having more inclusive data sets. You know, quite frankly, making better AI.

GUEST ZARAH GUILLEMET [00:37:08] Yeah, thank you, Nicol. I want to echo exactly what you just said about Zarah. She has been such a wonderful, impressive collaborator. I mean, as a high school student doing this level of research, it's really, really impressive. And I'm so proud of the work that she's done. I'm really looking forward to what she does in the future, whether that's in engineering, in tech, whatever direction she decides to take, I'm really looking forward to that. And then to your question about policy. I echo what Zarah says, which is, we need increased AI literacy for sure, people to understand where are these systems being used, how are they being used? What is your ability to opt in or opt out of their use? I think that is so important. But I also think there's still a place, maybe not in today's environment, but there's still a place for regulation to play. And I don't know if that's going to happen now or in five or 10 years from now, when it might be too late, or if we're just going to kind of ride on

the coattails of the EU. But there is a place for us to at least create some baseline standard and expectations of AI systems, right, and how they impact different communities based on demographics. And I think there is potential to get buy-in across the political spectrum for this, right? Because... We're not just thinking about race and ethnicity, right? Across educational divides, across rural-urban divide, right, there is going to be differences in how these tools are being used, how it affects these different communities. And so we should all be interested in understanding those things and doing more research on that. Similarly, there needs to still be this investment in research. Right. We're seeing maybe there's going to be a defunding of the National Science Foundation and other research programs across the federal government. But there's still such a place to understand the context and nuance of AI systems. Right. It's not just about endlessly pursuing better performance, more accuracy. There needs to be also equal value placed on human impact, social impact and whether we like it or not, AI is here, but we don't have to accept the way that it's here, right? We want to understand how it's affecting us and then be able to make change to that. And so how does that look like for policy, right, allowing users to have a way to address harms that happen, right, and we see this in other contexts, right. You can report posts on social media, right for example, Right, we need similar types of mechanisms for AI systems. And then the other place I think we need to be careful of is particularly in law enforcement and surveillance and in military applications. Is certainly no way we're gonna stop the train at this point in those groups using AI, but we can think about what are reasonable limits and appropriate uses where it could be integrated there. Our research showed that Those were the three places where Black women really felt the most strongly about the potential misuse of AI for their community. And so how do we use it in a way that's safe, that understands that people have different ways of interacting with AI systems, but also this institutional context.

CO-HOST NICOL TURNER LEE [00:40:49] Yeah, I think that is where I sit with so much respect for the beginnings of this research. And Zarah, you said it. This is about respectable data that goes into these systems as often generated through the lived experiences of the populations. Thank you both for joining me to discuss your insights, Raj and Zarah. I really appreciate it.

GUEST ZARAH GUILLEMET [00:41:13] Thank you so much for having us!

CO-HOST NICOL TURNER LEE [00:41:16] Thank you so much again for having us. You can find more about Raj's work at the Trustworthy, Intelligent, and Explainable

Robotics Lab or Tier Labs website at tierlab.commonsgc.cuny.edu. Again, tierlab.commonsgc.cuny.edu Please explore more in-depth content of tech policy issues like this at the TechTank newsletter that is on the Brookings website, accessible at Brookings.edu. Be sure to follow this conversation because there is going to be more at the AI Equity Lab where we try to bring conversations with different people, different industry sectors and different disciplines together to come up with what I call purposeful and pragmatic AI. Your feedback matters to us on the substance of this episode, so leave us a comment, share it with someone else, and listen to our future episodes, because you know us at TechTank. We are always making these issues more explainable and provoking new thought. This concludes another episode of the TechTank podcast. I'm Dr. Nicol Turner Lee, where we make bits into pound to bytes. Until next time, thank you for listening. Thank you for listening to TechTank, a series of roundtable discussions and interviews with technology experts and policymakers. For more conversations like this, subscribe to the podcast and sign up to receive the TechTank newsletter for more research and analysis from the Center for Technology Innovation at Brookings.