accessible video transcript

ON WHAT MATTERS

Making AI systems more just with Hilary Pennington and Dr. Timnit Gebru

This video transcript captures a Zoom conversation between Dr. Timnit Gebru, founder and executive director of DAIR, and Hilary Pennington, executive vice president of programs at the Ford Foundation.

To watch more videos in this series, go to Social Justice Leaders On What Matters and scroll down to access more videos in the series.

Transcript begins.

HILARY PENNINGTON: Hi, I'm Hilary Pennington. Thank you for joining us for the fifth conversation in our live series, “On What Matters.” I'm executive vice president of programs for the Ford Foundation. I'm a white woman, middle-aged with short hair, sitting in a brownish dress against a white wall. And I'm really excited today to talk to Dr. Timnit Gebru, who is an expert on artificial intelligence and technology, and especially on how to reduce the harm and the uneven benefits that it brings to society. And Timnit, over to you, to introduce yourself.

DR. TIMNIT GEBRU: Hi everyone. Thank you for joining us. My name is Timnit Gebru, and I am founder and executive director of DAIR, which stands for the Distributed AI Research Institute. I am a light-skinned Black woman sitting in front of a white wall background, and I'm wearing a maroon sweater.

HILARY PENNINGTON: So we're going to start by showing a video that just frames the issue that we're talking about, and, in particular, talks a little bit about the potential harms of algorithmic bias. So here's to the video.

> [Kade Crockford, Director, Tech for Liberty Program, ACLU Massachusetts. A white gender nonconforming person wearing business clothing.]

KADE CROCKFORD: Digital technologies, the information age, have changed the way that we live in ways that are really obvious—like, the fact that we all carry tracking devices with us everywhere we go—and in ways that are really opaque—like the various black box algorithms that every single day make decisions for us.

> [An animated 3-D black box spins as chains made up of ones and zeros flow into it on all sides.]

A black box algorithm is a mathematical formula that companies like Google and Facebook, as well as even governments, use to process large quantities of information and make decisions about what you'll see when you open up your web browser. They determine what price an airline will try to sell you a plane ticket for, and they can even determine how much your car loan will cost. That matters because it may very well be the case that someone in a rich, white neighborhood gets charged substantially less for auto lending than someone who lives in a largely poor, Black neighborhood—despite the fact that those people have pretty much identical driving records. This also happens in the employment context, where employers are using black box algorithms to sort through large quantities of data about applicants.

> [An animated 3-D black box floats above three separate stacks of resumes, scanning each stack with a laser beam.]

The algorithm will automatically sort and dispose of many, many applicants before any human being even enters the process to decide who's going to get the job or who will get an interview. And those types of systems are in use in almost every industry today. Right now, there's a major information asymmetry, right, between folks who work at Google and Facebook, about exactly what these tools are capable of, and what they're currently doing, and the vast majority of the public. We need to bridge that gap. And we need technologists alongside us in that fight. Fifty years ago, when there was no public interest pathway for law students, really, besides working for the ACLU, we were not doing all we could as a society, frankly, to maximize what it means to be a lawyer, to maximize the benefits of a legal education as far as, you know, impacting the society in general in a positive way. It's equally important now for technologists to also come to the table and tell lawmakers exactly what these tools are doing, you know, what the future looks like, and how to ensure that we don't, you know, magnify exponentially the existing inequalities in our society. If we don't bring those technologists into the public interest fold, I think we're really looking at a very dangerous world in which technology does exacerbate and exponentially increase those inequalities.

> [This is tech at work for the public! Hashtag Public Interest Tech. Ford Foundation dot org forward slash tech. Ford Foundation logo: a globe made up of a series of small, varied circles.]

HILARY PENNINGTON: So Timnit, let's get started. At Google, where you led cutting-edge research on the ethical implications of AI, can you talk a little bit about how your experience there, and your scholarship, led you to leave Google and made you decide to start DAIR?

DR. TIMNIT GEBRU: I, famously, also got fired from Google and, you know, when I started working at Google, I was co-leading a small research team called the “Ethical AI Team,” with my then co-lead Margaret Mitchell. She had actually started the team. So I had been hired, to my knowledge, to do exactly what I did when I got fired, which is alert the research world and the public as to the harms of some of the AI systems and mitigate their harms. So at the end, I wrote a paper called “Stochastic Parrots” on the dangers of large language models, which led to my firing, and that really clarified that I couldn't really do that kind of work in a setting like Google. And so I started a nonprofit called the Distributed AI Research Institute to do this work.

HILARY PENNINGTON: Well, it's really important work that you're doing there. And I think–let's talk a little bit about why you've called it that. You know, I think for all of us who are reading the newspaper these days or watching anything on social media, we have heard about ChatGPT, and also just the headlines about the potential harms of technology. So talk to us a little bit about why you've structured DAIR the way you have. Why is it called the Distributed AI Research Institute, and why is it so important to you to put people at the center of the work and the research?

DR. TIMNIT GEBRU: Yeah, it's really interesting because the first word that came to my mind when I wanted to–when I was thinking about starting a research institute was “distributed,” and I actually called Eric Sears of MacArthur Foundation, and I was like, “Does this sound fine to you?” You might, you know, might sound wild, but I really want an institute where there are people distributed around the world, who don't have to leave their communities, that are embedded in their own communities, impacting the trajectory of AI development, right? I think that that is really a counterpoint to the centralization of power that we are currently seeing right now. So you mentioned ChatGPT, right? Even if the claims by these companies that, for instance, things like ChatGPT could replace lawyers or doctors were true, which I don't believe are true, what they would want is, you replace all the lawyers in the world and all the doctors in the world, and pay one company located in one location for all the services that you want in the world. What I want is something that is the complete opposite. Have a distributed group of people around the world, impacting tech and other societal development.

HILARY PENNINGTON: Well, I love that you talk about and you imagine not only what you don't want but what you want from technology. And so, let's dig a little bit deeper into some of the concerns that have animated you. And can you talk some more about some of your ethical concerns about AI? And, you know, specifically, its impact on the environment? I think in a way, most people don't know that it's an incredibly carbon intensive industry but more broadly than that.

DR. TIMNIT GEBRU: So, you know, you mentioned, for example, the news on ChatGPT, right? And some of the news is people speculating about whether these systems are sentient or they feel things. And I think the reason that this happens is because there is an intentional obfuscation of all of the things that are involved in creating these systems, whether it is exploited workers who are providing data and labeling data and suffering from PTSD while looking at toxic words and images, or the amount of compute power that is required to train some of these models, right? We often talk about things like, we say the cloud, where, you know, things are not actually being processed in a cloud. They're being processed in large data centers that take lots of water, for instance, from local communities and that require lots of energy. And even if it is, quote unquote, “from carbon neutral sources,” which some companies claim, it's not free, right? There are resources that are needed—minerals that are needed to create some of these even renewable energy sources. Trees that are cut to build these data centers. So our concern, especially with large language model based systems like ChatGPT, when I wrote even my paper, was that the people who are benefiting from these kinds of systems, like, for instance, predominantly English-speaking people, let's say living in the U.S., are not necessarily the people paying the costs of the ensuing climate catastrophe because of the energy consumption of these models, whether it is within countries like the U.S., for instance, or across countries.

HILARY PENNINGTON: Timnit, just for a minute, can you explain how AI actually is made?

DR. TIMNIT GEBRU: This is a really good question because, how the sausage is made, as people say, is really an important aspect of actually some of the issues that are important to put forth as well. And for the audience who might be interested, we wrote a paper called–actually an article called, “The Exploited Workers Behind AI,” just so that people know how many people are involved in this process. So many–most of the systems that people call “AI” right now, are based on what we call machine learning. And these types of systems, and specifically deep learning–I don't want to get into a lot of jargon–but the important point to note is that many of these systems require a lot of data and a lot of compute. So you first have input data–many times that is labeled according to a particular type. So, for instance, let's say what you want to do is classify whether there is a cat in a photo, right? So you, many times, need lots and lots of photos with cats and maybe without cats, and people to label whether there are cats in those photos, or not. And of course, people that supply those photos as well, right? And then, these models are trained on lots of those kinds of photos. And finally, you have a model that has sort of learned how to classify a new photo with a cat or not with a cat. This is just kind of like a simplified example. Now, at all ends of those spectrum, you have a lot of humans involved, right? Humans supplying the data, oftentimes, without their consent, especially nowadays, we're seeing artists fighting back. We have humans labeling the data. Oftentimes, if the data is actually toxic or disturbing, you have–similar to content moderators–you have these humans sifting through lots and lots of data and having to see all of this horrible content, giving them PTSD. And you have, of course, the compute that, as we talked about, has a lot of environmental footprint because we have all these data centers that are required to train many of these models.

HILARY PENNINGTON: Let's go back for a minute to one of the things you said. You talked about language models and the predominance of the English language being used to build these models. And you also talked about the word sentient, as in feelings. Do these, you know, does the model have a feeling? Talk a little bit more about the ways in which these systems get constructed, and why and how it is that they cannot actually be sentient.

DR. TIMNIT GEBRU: So we actually dubbed the term “stochastic parrots.” Actually, I say we—I can't take credit for it—it was my colleague, Emily Bender. And what that means, you know, we're trying to show that when you see a parrot speaking, supposedly coherently, grammatically correct sentences, you don't think that it's extremely intelligent or that it's understanding what it's saying, but you think, “oh, it's cute,” right? And so these systems are, in a similar way, have learned to stitch together words to create the most likely sequences of words, according to what they see on the internet. So they've been trained on lots and lots of data and text from the internet and have been trained to, again, output the most likely sequences of words. So when we see the outputs–these outputs–we as humans can attribute intent to these outputs and feel like it might be coming from another human and forget things that are even as complicated as large language models, right? In the sixties, there was a system called Eliza, a chatbot called Eliza that was much, much simpler, right? This is in the sixties, and people still talked about it like it was some sort of person, and its creator actually was very distressed by the ethical concerns raised by such systems, even at that time.

HILARY PENNINGTON: Yeah, even at that time. Well, you know, sometimes when you look at what we're up against, right, and you look at the black box, as it was referred to in that video, you look at the size of the companies that are developing the AI, it’s easy to feel that even you, and even DAIR, and all the other amazing and smart activists are a little bit like David up against Goliath. You know, it just seems so big–the forces that you are fighting against. And yet, on the other hand, of course, we know, David won. David defeated Goliath. So tell me a little bit about some of the progress that you've made at DAIR that you are proud and excited about.

DR. TIMNIT GEBRU: So I think that, yes, sometimes it does feel that way. I mean, we don't have the $10 billion that OpenAI just received from Microsoft, but I have to remember that it is people. A collective power is the biggest form of power, right? And when society decides that they don't want something or they've had enough, they've been able to move forward in the past. So that's really what I draw inspiration from. And for DAIR, I think the fact–just even our continued existence–and that I find that most of my team members really want this organization to exist, that's our first win. You know, we have not only researchers in computer science, and sociologists, and engineers, but labor organizers and refugee advocates working together to chart the future of this technology. And we've been able to, already, work on projects that actually put data in the hands of people in marginalized groups. So one of our early projects that we're still working on is on understanding the continued legacy of spatial apartheid. So this is the legacy of apartheid South Africa, right? And so we use computer vision techniques and satellite images to draw the boundaries of townships so that people in townships can actually have firm evidence that their quality of life is not as good as those of suburbs that were delineated during apartheid. And this work was also led by Raesetje Sefala, who is someone who grew up–who was born and raised in a township. So to me, these kinds of projects are–they're what give me hope and they're what, I think, you know, move me forward in believing in people's power and people's collective power.

HILARY PENNINGTON: That the people are the stone that fells the giant. I love that. Well, we've gotten a lot of questions from the audience, and I'm going to turn to some of those now. And I'm going to start with a question from Crystal, who wants to know, “What are the biggest factors making AI regulation complicated? What will privacy look like with AI and what will machine learning in the private sector and government look like?”

DR. TIMNIT GEBRU: That's a really good question. There are a number of factors complicating AI regulation. One of them is that regulators feel like they're not well-equipped to regulate because they don't necessarily know how these systems work. But, you know, I always remind them, they don't have to know how these systems work in order to regulate. They only need to understand the impacts, right? So the people who know how to build something are not necessarily the best people who know how to evaluate societal impact. So that's one. The other one–the biggest one to me, I think, is the imbalance in resources. We're many times in a situation where the people who are harmed–the onus is put on them to show–to prove harm. Even in the cases where regulation is present, like, the general data regulation protection–the GDPR–in the EU, right? A lot of times the onus is on the individuals who are harmed to prove that they are harmed, rather than putting the onus on tech companies before they put out products, proving to us that they are not harmful. And finally, the agencies themselves that we expect to work on regulation are very understaffed, and they also have a hard time enforcing regulation, even once you have the actual regulation in place. And of course, we can't forget all the lobbying that–speaking of imbalance of power and resources–that biotech companies have to sway the interest in their favor.

HILARY PENNINGTON: Well, what you describe is really sobering, too, because even if we had better regulation, how can it be enforced? What do you think about that?

DR. TIMNIT GEBRU: I think honestly, it's really about, kind of, we need to have even regulation that we have right now–that I'm sure is being broken–like worker exploitation, or union busting, right? That could do a lot if we really enforced and punished companies who break these laws. That can move us forward, even when we're talking about AI. But the problem is, again, there is a lot of lobbying, right? And then, there is also, you have agencies that are understaffed and under-resourced, where we are expecting to go up against these companies. So we need to–at even the agencies that we have right now and the laws that we have right now–we need to ensure that they have adequate resources and staff to enforce existing laws. And we need the punishment to have teeth, so that it actually forms as a deterrent.

HILARY PENNINGTON: Yes. All right. Well, let's go to the second question, which is from Monica, who says, “Data can be biased because people are biased. How will the inputs to AI be, quote ‘corrected,’ as to avoid biased assessments, recommendations, or actions? And who is deciding what ‘corrections’ will be made? How will we ever reach an equitable application of this technology?”

DR. TIMNIT GEBRU: This is such a great question because many people in the field of AI want to make people think that there is such a thing as a neutral technology. In fact, Sam Altman, who is the CEO of OpenAI, even tweeted saying that ChatGPT should not have any political inclinations or should be, quote unquote, “neutral”. And a whole bunch of people were like, “What about, you know, so fascism will be okay? Or Nazism?” You know, you can't be neutral. There is no such thing as being neutral. And this question, I think, gets to that, right? There is no such thing as a neutral or unbiased data set. So, what we need to do is make those biases and make those values that are encoded in it, clear. I wrote a paper called “Lessons from Archives” with a historian, whose name is Unso Jo, and we were talking about how archivists, right, they have curators when they're curating data, and we know that the curator’s point of view is going to be encoded in there. Whereas in this case, people think that just because they're putting data from the internet, that everybody's point of view is represented. And that's just not true, right? So what we need to do is make those values explicit, and those values are often decided by society. We have laws already, existing laws, saying what's okay and what's not okay. So I think that's what I want people to know is, like, there is no such thing as a completely unbiased data set.

HILARY PENNINGTON: Well, and I love the emphasis you're putting on making it transparent, making it visible. So I'm going to go to our last question, which comes from Erin, and that's turning to action that all of us could take. So Erin says, “How should those of us with influence in the corporate social impact space be steering our companies in relation to AI, especially technology companies whose philanthropic strategy is tied to making a charitable digital impact? Where should we be putting our efforts?”

DR. TIMNIT GEBRU: So this is a great question. I think that if you're at a company that is trying to invest in philanthropic ventures, I would say that you should ensure that they're not doing it in order to then go back and skirt regulation and show, see, these are all the places that we're investing in. And so you shouldn't regulate us. So I think that's really important to make sure that they're investing in grassroots organizations and other organizations. But we need to make sure that the net impact is positive, rather than resulting in not regulating, holding them accountable. The second thing I would say is that, if you're inside a company, you know, you can always–a collective action is really important–you can always have partnerships with people outside of the organization and kind of funnel information because often people inside companies know what's going on before the rest of us do. And third, it's really important to move resources to people who are very much impacted by this technology and know firsthand what is needed to counter these negative impacts. But they're not legible to, let's say, philanthropists, they're not legible to funders, so they don't get all these resources, right? And so it's really important to build connections and networks with groups of people that you generally wouldn't see in your day-to-day. Find those people, follow them, whether it's on social media or conferences or wherever. You might find people like that. Build relationships and personal connections and invest in those people and organizations. So for instance, of course, DAIR–I have to plug my own organization–but there's organizations like Mijente that I love, and there's organizations like AGL, AINow, Data and Society, etc. And so, yeah, this is my recommendation.

HILARY PENNINGTON: I love those examples. They give me hope. And I want to close with just a final question, which I always ask, which has to do with what gives you hope, especially given such a challenging area that you work on, with so much change, and so much uncertainty, and so many aspects of it, that are frankly, really scary. What gives you hope? What keeps you going?

DR. TIMNIT GEBRU: I think it's going back to what I said earlier, that even though there's all this money and centralization of power, I always try to remember that it is collective action and people power that is the biggest form of power. And so, oftentimes, when I see the next generation, when I talk to students, when I talk to other people organizing, I see how passionate they are, and how optimistic they are, about building a different path and a better future. So these are the things that give me hope, talking to these people and seeing the next generation.

HILARY PENNINGTON: Thank you. Thank you so much for just a wonderful conversation and thank you to everyone who has joined us to listen to it. And please be ready to join us for the next conversation in the series, which is with Saliem Fakir. And stay tuned for some details about that. Thank you so much, Timnit.

DR. TIMNIT GEBRU: Thank you.

End of transcript.