Ex-Google AI skeptic Timnit Gebru starts anew

March 26 2022, by Matt O'brien

Timnit Gebru poses for photos in Stanford, Calif., Monday, March 21, 2022. When she co-led Google's Ethical AI team, Gebru was a prominent insider voice questioning the tech industry's approach to artificial intelligence. Credit: AP Photo/Jeff Chiu

When she co-led Google's Ethical AI team, Timnit Gebru was a

prominent insider voice questioning the tech industry's approach to artificial intelligence.

That was before Google pushed her out of the company more than a year ago. Now Gebru is trying to make change from the outside as the founder of the Distributed Artificial Intelligence Research Institute, or DAIR.

Born to Eritrean parents in Ethiopia, Gebru spoke with The Associated Press recently about how poorly Big Tech's AI priorities—and its AI-fueled social media platforms—serve Africa and elsewhere. The new institute focuses on AI research from the perspective of the places and people most likely to experience its harms.

She's also co-founder of the group Black in AI, which promotes Black employment and leadership in the field. And she's known for co-authoring a landmark 2018 study that found racial and gender bias in facial recognition software. The interview has been edited for length and clarity.

Q: What was the impetus for DAIR?

A: After I got fired from Google, I knew I'd be blacklisted from a whole bunch of large tech companies. The ones that I wouldn't be—it would be just very difficult to work in that kind of environment. I just wasn't going to do that anymore. When I decided to (start DAIR), the very first thing that came to my mind is that I want it to be distributed. I saw how people in certain places just can't influence the actions of tech companies and the course that AI development is taking. If there is AI to be built or researched, how do you do it well? You want to involve communities that are usually at the margins so that they can benefit. When there's cases when it should not be built, we can say, 'Well, this should not be built.' We're not coming at it from a perspective of tech

solutionism.

Q: What are the most concerning AI applications that deserve more scrutiny?

A: What's so depressing to me is that even applications where now so many people seem to be more aware about the harms—they are increasing rather than decreasing. We've been talking about face recognition and surveillance based on this technology for a long time. There are some wins: a number of cities and municipalities have banned the use of facial recognition by law enforcement, for instance. But then the government is using all of these technologies that we've been warning about. First, in warfare, and then to keep the refugees—as a result of that warfare—out. So at the U.S.-Mexico border, you'll see all sorts of automated things that you haven't seen before. The number one way in which we're using this technology is to keep people out.

Q: Can you describe some of the projects DAIR is pursuing that might not have happened elsewhere?

A: One of the things we're focused on is the process by which we do this research. One of our initial projects is about using satellite imagery to study spatial apartheid in South Africa. Our research fellow (Raesetje Sefala) is someone who grew up in a township. It's not her studying some other community and swooping in. It's her doing things that are relevant to her community. We're working on visualizations to figure out how to communicate our results to the general public. We're thinking carefully about who do we want to reach.

Q: Why the emphasis on distribution?

A: Technology affects the entire world right now and there's a huge imbalance between those who are producing it and influencing its

development, and those who are are feeling the harms. Talking about the African continent, it's paying a huge cost for climate change that it didn't cause. And then we're using AI technology to keep out climate refugees. It's just a double punishment, right? In order to reverse that, I think we need to make sure that we advocate for the people who are not at the table, who are not driving this development and influencing its future, to be able to have the opportunity to do that.

Q: What got you interested in AI and computer vision?

A: I did not make the connection between being an engineer or a scientist and, you know, wars or labor issues or anything like that. For a big part of my life, I was just thinking about what subjects I liked. I was interested in circuit design. And then I also liked music. I played piano for a long time and so I wanted to combine a number of my interests together. And then I found the audio group at Apple. And then when I was coming back to doing a master's and Ph.D., I took a class on image processing that touched on computer vision.

Q: How has your Google experience changed your approach?

A: When I was at Google, I spent so much of my time trying to change people's behavior. For instance, they would organize a workshop and they would have all men—like 15 of them—and I would just send them an email, 'Look, you can't just have a workshop like that.' I'm now spending more of my energy thinking about what I want to build and how to support the people who are already on the right side of an issue. I can't be spending all of my time just trying to reform other people. There's plenty of people who want to do things differently, but just aren't in a position of power to do that.

Q: Do you think what happened to you at Google has brought more scrutiny to some of the concerns you had about language learning

models? Could you describe what they are?

Q: Part of what happened to me at Google was related to a paper we wrote about large language models—a type of language technology. Google search uses it to rank queries or those question-and-answer boxes that you see, machine translation, autocorrect and a whole bunch of other stuff. And we were seeing this rush to adopt larger and larger language models with more data, more compute power, and we wanted to warn people against that rush and to think about the potential negative consequences. I don't think the paper would have made waves if they didn't fire me. I am happy that it brought attention to this issue. I think that it would have been hard to get people to think about large language models if it wasn't for this. I mean, I wish I didn't get fired, obviously.

Q: In the U.S., are there actions that you're looking for from the White House and Congress to reduce some of AI's potential harms?

A: Right now there's just no regulation. I'd like for some sort of law such that tech companies have to prove to us that they're not causing harms. Every time they introduce a new technology, the onus is on the citizens to prove that something is harmful, and even then we have to fight to be heard. Many years later there might be talk about regulation—then the tech companies have moved on to the next thing. That's not how drug companies operate. They wouldn't be rewarded for not looking (into potential harms)—they'd be punished for not looking. We need to have that kind of standard for tech companies.