Machine Learning Fairness in Big Tech

Auriel Wright talks about her work on advancing fairness and equity in computer vision at Google.

By Adinawa Adjagbodjou

`OPEN ACCESS`

Auriel Wright is a product manager at Google Research and a product mentor (PM) at Google for Startups. Before graduating with a computer science degree from Harvard College, she founded her own startup, StrattyX, through Harvard's Innovation Lab. Beyond her core role as a machine learning PM, Wright is a mentor for various startups and dedicates her time to supporting young tech professionals and entrepreneurs through their journeys, teaching CS workshops for several universities, orchestrating founder dinners, and connecting CS majors through happy hours.

Adinawa Adjagbodjou [AA]: Can you tell me more about what you do in your role?

Auriel Wright [AW]: So, I work on machine learning fairness problems inside of Google. I help build metrics to think about what is sociologically fair to think about in computer vision. After we decide what metric is considered fair, I spend time actually putting that in practice inside of our models and making sure our models adhere to these fairness metrics. I also check that they're being upheld in the long term. It's important to look for unfair biases and ensure that when we build the model, it's created with fairness and inclusion in mind. This is my job at Google. This is all I do. I think about fairness and machine learning [ML] signals all day.

AA: That is really, really interesting. I'm curious if there's work that you did before that led you to wanting to do this.

AW: Actually, yes. I wrote my capstone research paper on something related. I took a course at Harvard for my computer science degree that resulted in a research paper. For my paper I designed this program for detecting my heart rate using my own video signal. It's optimized for my darker skin tone so that I could measure my heart rate. Through this research project, I was kind of drawn into fairness in ML and the work I do now.

I was originally just planning on using established methods to do the analysis on saved videos instead of streams of video. But I realized that with the existing papers online, the models that they suggested weren't working for my skin tone. But they worked for my white boyfriend. He'd sit in front of my own project made from theories other people suggested I use and it would project him perfectly, but not me—its creator! And I thought, "I literally built this! How can it not work for me?" So I went back to the drawing board. I read other papers and later, I was able to optimize it through trial and error to work for my skin tone. And a spoiler alert—if you ever decide to recreate a project like this—what I saw was that my darker skin tone benefits from looking at my undertones of green, not red.

The core underlying problem was I was trying to use methodology made to track white people's vital signs and I was not white!

So yes, this kind of got me interested and I decided to start working on similar projects at Google.

AA: Can you expand on that last statement and what it means on a day-to-day basis and on a larger scale within the company and the space?

AW: Google's AI principles were established and enforced starting a few years ago.

Black engineers have been in the minority, and sometimes product development would not always factor in the way that they might experience technology. There's already bias in the selection of who the product is going to be tested on. There were times where people of my skin tone just never got an opportunity to test developing products. And next thing you know, a product does not work for me because people like me were not in the beta test. Like, many teams don't have an Auriel sitting around. Well, I guess if you work at Google you can technically have access to one, but there's no telling if I have the time to give you and every other product at Google attention for your visual perception problems.

This is a pipeline issue. To be able to identify and address some inherent biases when it comes to building a product, the thing that I think is most effective is trying to build systematic solutions to these really systemic problems. Like, Black people are not that common in Big Tech because Black people were also horribly mistreated for the last two centuries and denied economic equality, which now means Black people are not at the table to have these discussions because they didn't even have chairs to begin with.

So there needs to be a system in place. Probably a mixture of sociological, economic, and technological solutions to be able to make sure that even if we can't exactly undo the biases, you can at least put safeguards in place to make sure that the biases are not perpetuated.

AA: Definitely. So I want to know if there's ways that you want to see change. You mentioned representative data to make sure things work for everyone for example. Are there ways that you think the envelope can be pushed further when it comes to the work being done?

AW: Well, there are so many things, but I'm going to say three and try to prioritize them.

Step one, I think product leaders should get the goals and functionalities from user experience design (UXD) and research (UXR) for the product in mind, and work backwards from there to figure out what data they need to be able to make sure that that's going to happen. Let's say the goal is for the computer vision model

to perceive if an image includes a person. If the definition of a perception model is to literally, in the computer vision context, to be able to see the person and you can't see the person, then make sure you clear with UXD what "person" means. Then work with UXR to figure out how to get the model to see each "person." This is the strategy of making it "work" at the most basic level.

Another thing you should do is to make sure that you write out and you review with a diverse group of people what you want the product to do, and then work backwards from there. Like, don't try to just build the minimum viable product of some particular thing in computer vision. It is one of those things where you should be a little more systematic and think about it. Because the harm that you can do when it comes to not

seeing a person is huge. Imagine having a whole group of people—who make up a third of the population—who you can't see. That just does not sound like it works.

So, the second one is to admit to yourself that biases happen and move on. Don't be a person that just talks about it—do something about it. I care about your actions. There are a good number of people who throw around the word fairness in AI. I ask, what do you define as fair and how did you get to that point tangibly? Is there any rigor, any research at all around it? And when you feel like you've gotten to the point where you even think fairness exists, what caveats do you have or what do you think we have left to be able to get to equity? Because being fair and being equitable— they're oftentimes used interchangeably, but just because something is fair does not mean that it's fully equitable. The fairness and equity discussion has to be its own post that maybe I'll write one day.

And then last but not least, lean on your user experience researchers. I feel like machine learning researchers tend to get in their head about the process of building their model, or they seem to want to have, like, the perfect layers to be able to map everything perfectly. They want a specific input that a rater gave them to perfectly map out to an output for their work. So they'll do tons of distillation, go back and try different regressions and all these different loss functions and get into the nitty gritty and weeds of trying to map the answer to the problem.

EDITORIAL NOTE

The views and opinions expressed in this interview belong to the interviewee and do not necessarily represent those of Google.

Biography

Adinawa Adjagbodjou is a Ph.D. student in human-computer interaction at Carnegie Mellon University focusing on the design of digital and immersive technology in partnership with marginalized communities. Adjagbodjou's most recent research has focused on designing virtual reality environments to promote collaboration, equity, and creativity, as well as speculative design with Black community organizations and practitioners.