

JASON NEWTON Data, and the algorithms that interpret data, are everywhere. From predicting which advertisements to send to your social media feeds, to using your location data to track which stores you frequent, society constantly leverages data to make decisions – even in the criminal justice system. Data and algorithms are being used in policing to predict where crimes might take place, and during sentencing to predict whether or not someone convicted of a crime is likely to reoffend.

But with the long history of systemic racism within the United States criminal justice system, does that mean the algorithm itself will be flawed by racist and prejudicial assumptions?

JN I'm Jason Newton...

RACHEL TRAUGHBER ...and I'm Rachel Traughber, and this is Unequal: a Harvard University series exploring race and inequality across the United States.

RT To answer this question and shed light on how our data and algorithms are being used more broadly, we're joined today by computer scientist Latanya Sweeney, Professor of the Practice of Government and Technology at the Harvard Kennedy School and in the Harvard Faculty of Arts and Sciences. She is an expert on data privacy, director of the data privacy laboratory at Harvard, and is widely regarded as one of the most preeminent voices in the field. She joins us now with more.

JN Professor Sweeney, give us a brief description of what these data algorithms are, what they tell us and how is it they are now used in our criminal justice system.

LATANYA SWEENEY Sure, thank you. Society has definitely experienced exponential growth in the amount of information collected on individuals. The idea now is not only do you have the collected data, but how do we use that data to make decisions. And that's where these data algorithms come in. They analyze the data, they look for models, they make predictions, and they can be used to help humans make decisions.

One of the places they can do that is in the criminal justice system. There are all types of decisions that are made there and lots of historical data on which an algorithm can learn. One of those examples is recidivism: This is the place where you try to decide whether or not a person will be allowed to go out on bail. So just because someone's arrested doesn't mean they're guilty, that's where the trial to decide. But whether or not they'll be allowed to go out on bail is a decision that has to be made, usually by a judge, and a recidivism algorithm can use historical data to make that recommendation.

JN Okay, but could you tell us what some of the drawbacks are to using this technology, as it tries to predict human behavior?

LS Definitely, there are two major areas in which these drawbacks can happen. One is the date on which the algorithm has to learn or it's going to learn these models to make its predictions, and the other one is when the algorithm uses a feedback from the person who is giving the information. And they say, Yes, I like that it gives me more like that, or give me less like that. So the bias in the algorithm example, is in the recidivism algorithm that I talked about, if judges before had been making biased decisions about who was allowed to go out on bail and who wasn't, the algorithm will pick up that bias, and will then continue to perpetuate those decisions. If the algorithm has a feedback system, and it makes a recommendation, and a judge doesn't

like it says, I don't really agree with that one, then over time, the algorithm will adjust to the preferences or biases that the judge may have.

The example I think that most people really understand is when people are being recommended for a job. So there are many algorithms now that will take a bank of resumes and when there's a job opening, or recommend people based on their resume to the potential employer. The potential employer then says yes, this one comes for an interview, not those. But if that employer has a bias against older people, then in fact, only young people will come be offered to offered to that person for interview. And on the next time that employee use it, it will only recommend younger people. So now the bias is even trapped in even further.

RT I would like to unpack a little bit about these data algorithms applied to search results. I know that you've done some particular research on this and had some interesting discoveries. And I wonder if you would share a little bit about that with us?

LS: In fact, that's what started this area of research on algorithmic fairness, I had just become a professor here at Harvard. And I was in my office, when I was being interviewed by a reporter. I wanted to show him a particular paper. So I typed My name into the search bar onto the Google search bar, and up pop the paper I was looking for, but also a popped an ad implying I had an arrest record. And so the reporter said, I forget that paper, tell me about when you're arrested. And I said, Well, I haven't been arrested. But eventually I had to click on the link, pay the money, and show him that not only did it not have an arrest record for me, but with my unusual name, there was no one else who had an arrest record with that name, either.

But that started me on the question of why did that happen. And so I spent hours and then a couple of months, and what I came to learn was that those ads came up, implying you had an arrest record, if you search the name of a real person, their first and last name, and if their first name was given more often to a black baby than a white baby, and they were 80% likely to have an arrest record show up. And and the opposite. If the baby the first name was given more often to white babies.

Discrimination in the United States is not illegal. But [for] certain people in certain situations, discrimination is illegal. One of those...one of those groups of people are Blacks, and one of those situations is employment. And so if two people were applying for a job, and a search implies that one of them has an arrest record and the other one doesn't, they're put at a disadvantage. So this was the first time an algorithm, in this case, Google Search, was found to be in violation of the Civil Rights Act.

JN That's a really good point. So from the bias in the in the courtroom to out in the community, brings me to my next question, how have aggressive police tactics, and the over policing of black and minority neighborhoods over the past several decades affected this type of algorithm?

LS Right, this is a fantastic question. We don't we can't even say we know the exact answer because we have to be able to say "this is exactly its impact on the data." As we look at studies who show its impact on the data, then the question is, can we account for that when the algorithms are learning? So if the data is biased, the result from the algorithm is going to be biased?

I'll give you a Harvard example. Many years ago, when I was at Harvard, you know, pretty much it was primarily young white men, or certainly we can imagine a time even earlier than that, where was young white men. If we wanted to make an admissions algorithm for Harvard at that time, we would have used the admissions data that had that had been used. The algorithm would learn things and the population that we would get would be young white guys. So the data we provide to the algorithm has a lot to determine about what the algorithm will put out. Often I speak to students, and it really brings that home to them. Harvard's campus right now is so diverse. When you look at in the classroom, it's just an amazing, these are the best minds in the world. And they come from all different walks of life, all different countries. When they say wait a second, this room would look - look how different it would look, and not based on the same criteria.

RT We've spoken a lot about how these algorithms can be used nefariously or unwisely, I'm wondering if you can suggest some opportunities that they could be adjusted for good and what that future might look like.

LS Well, I'm a computer scientist, I want society to enjoy the benefits of these new technologies! I want society to enjoy the benefits, but without sacrificing these important historical protections. There's no reason that that can't be done. In every example that I've talked about today, we can imagine and envision ways two different things that can be done. One is an analysis of the data and the algorithm that can provide a guarantee of what it won't do, of what kind of bias it doesn't exhibit. The second kind of thing we can imagine is actually changing the algorithm so that it itself does it offsets a racial bias result. So the advertising algorithm would be an example of that.

RT And last, I just wonder if you might give us any insight into you know, what kind of oversight goes into algorithms in general? I mean, it sounds like based on the conversation we've had here today that people are sort of, this is new technology, people are just applying it as they see fit, because they think it might make their jobs easier, or their lives easier. And instead, it's...it's exacerbating some of the inequalities that we're seeing. Is there any oversight of these from a governmental perspective from a nonprofit perspective? Like, how, what is the next step in terms of policy creation?

LS Yeah, this is a fantastic question. You know, it's part of a larger arc. It's not just these algorithms, it's technology design, and designers are the new policymakers. We don't elect them, we don't vote for them, but the arbitrary decisions they make, and the technology products they produce dictate how we live our lives. And that's everything from free speech, to...to privacy, to things like these decision-making algorithms that we've been talking about today. And so there's a bigger arc around, how do we address technology itself and its impact on society. Our historical protections are intact, but they don't seem to know how to apply or adapt themselves to the current technology.

So one big answer we need is a new kind of technologist, one who works in the public interest, working in all of these places, and [with] all these stakeholders, working in technology design and technology companies working in agencies that regulate for these rules, so that they can understand them, working in Congress, and so forth.

RT Yeah, I think we've definitely seen what happens when we don't have people doing that work. But the you know, the events of January 6, in particular, and how a lot of that liaising to make that experience happen happened online happened through technology not being

regulated. So it would be interesting to see what the next step would be to prevent something like that, and how technology could be used productively and positively to help.

LS Yeah, and I would just say, you know, Twitter provides a great example, as well. Twitter has its own definition of free speech. And it's a private company, so it can have this other definition of free speech. But Americans want to talk about Americans definition, America's definition of free speech, and somehow be upset if Twitter doesn't quite satisfy it. And and the other thing I've heard is Americans interpreting Twitter's version as if it is America's version. It just shows you how pervasive the technology is, and how the design of Twitter as an example, is really dictating what we think our free speech rights are.

RT If you liked what you heard today and are anxious to learn more from Harvard's Unequal project, visit us online at the "in focus" section of harvard.edu.