TRTakademi
ISSN 2149-9446 | Cilt 06 | Sayı 11 | Ocak 2021 | Büyük Veri

Data, Algorithms, Fairness, Accountability

DANAH BOYD
Founder and President of Data & Society Research Institute

As members of the US Department of Commerce's Data Advisory Council, we are all deeply committed to seeing government use data strategically and productively to improve our country. We have spent the last year examining ways to open up government data and push for more engagement with government data. My talk today is intended to challenge and provoke us to think more deeply about the mission that we're committed to. I want to challenge some of the basic assumptions that we all hold dear and highlight how some of our values are in conflict. We all assume that our commitment to using data well is a commitment to using data for social good. But what if our passion project will increase inequality and hurt the people we're trying to help? What if our efforts will do harm?

How many of you think that discrimination is a bad thing?
This is a trick question because it all depends on how we define discrimination. When I say discrimination, most people think about unjust and prejudicial treatment based on protected categories. But discrimination as a concept has mathematical and economic roots that are core to data analysis. The practices of data cleaning, clustering data, running statistical correlations, etc. are all practices of using information to discern between one set of information and another. They are, in essence, a form of legitimate mathematical discrimination. The big question presented by data practices is: Who gets to choose what is acceptable discrimination? Who gets to choose what values and trade-offs are given priority?

There is nothing about doing data analysis that is neutral. What and how data is collected, how the data is cleaned and stored, what models are constructed, and what questions are asked - all of this is deeply political. Do not for a second pretend that we can build a neutral platform or punt the political implications of

data down the line. Every decision matters, including the decision to make data open and the decision to collect certain types of data and not others.

Open Data: Tool for Self-Segregation?

Let's begin by talking about open data, an issue that many of us in the room care deeply about. There's a gut instinct in open data communities that making data available to the public is a good thing. That it's democratiizing. But what if it's not?

In NYC, the Department of Education has opened up data about schools through the School Performance Dashboard. If you don't like their interface, you can look at Inside Schools, which is also mostly powered by NY DOE data. The purpose of these services is to help empower parents and families to make the best school choice for them and their family. But school choice is political. And the data that the DoE collects runs straight to the fraught nature of the value of education.

What makes a good school? Test scores? Student makeup? Parent ratings? Different families have different values so they read the data differently. This is considered a feature, not a bug, because what families want from schools differs.

Unfortunately, school choice based on data presents a series of challenges. First, there's the very real reality that data helps some families more than others. There's a huge variation in ability to read statistics, not to mention English. School ranking is connected to geography and some families have more mobility than others. Furthermore, some families have more time to devote to understanding what the variables mean in terms of the schools themselves. Or, if you're wealthy, there are actually expensive private services that will analyze the data for you and help you weigh your options. You get the idea. All of this showcases unevenness in being "informed" and the limits of "choice" that is not fixed by making data available.

What's not discussed is how public good and individual desire often conflict. Mahzarin Banaji has done fantastic work at Harvard highlighting how hard diversity is in the workforce. It doesn't matter that more diverse teams are more successful. They believe themselves to be less successful and they say that they are less happy. Given choice, workers self-segregate even if that's not beneficial for the company or for society.

Guess what? School choice prompts people to self-segregate for the exact same reasons. Black families choose schools that are predominantly black; white families choose schools that are predominantly white. They did this long before open data, but with the rise of open data, self-segregation has escalated. NYC schools

are now the MOST segregated schools in the country. Open data enabled people to segregate even though we know that this has serious long-term social and individual repercussions. Even privileged children are better off in diverse environments and, yet, most privileged parents will opt otherwise if given the choice.

When we open up data, are we empowering people to come together? Or to come apart? Who defines the values that we should be working towards? Who checks to make sure that's what our data projects are helping us achieve? If we aren't clear about what we want and the trade-offs that are involved, simply opening up data can - and often does - reify existing inequities and structural problems in society. Is that really what we're aiming to do?

Criminal Justice: Equity or Equality?

Unless you're a statistician, you probably haven't been following the debates around Northpointe's COMPAS, an algorithmic tool that is used to assess whether or not someone who has been arrested is a high risk to society. This information is used by judges to help determine if someone deserves to receive bail. If you don't know much about the US criminal justice system, bail is predictive of just about everything. If you get bail, you're more likely to keep your job, your house, your children, your spouse. If you don't get bail, you're more likely to plead guilty, even when you're not.

So how does a judge fairly determine if someone deserves bail? Historically, judicial decisions have been extraordinarily biased if not outright racist. "Risk assessment tools" have been developed in order to help neutralize analysis and help judges make better decisions, presuming judges with "neutral" third party information will be better at making informed decisions. But what's at stake is that not everyone agrees on what are acceptable outcomes, let alone acceptable trade-offs.

A few months ago, ProPublica published a controversial article arguing that there is nothing equitable about COMPAS, that it actually produces unfair outcomes for people of color and, most notably, blacks. (COI notice: one of the Data & Society fellows helped do the analysis.) They showed that blacks who never reoffended (one of the cornerstones of the algorithm) were twice as likely to be classified as medium or high risk than whites and, thus, be denied bail more often. Northpointe retorted by highlighting that they designed the system such that blacks and whites are equally likely to reoffend based on their score. Scholars weighed in, debates ensued. What becomes crystal clear is that there's no clear definition of legal fairness. And, more importantly, what's at stake comes down to a disagreement around false positives versus false negatives. Equality of likelihood versus

equity of resultant outcomes. Interestingly for the statisticians, there's no way to resolve the two different approaches to fairness which means someone is going to get screwed no matter what.

A huge part of the underlying problem stems from the limits of the data that are being used. Criminal justice data is extraordinarily biased. Black and brown people in the United States are more likely to be arrested for the same activities as whites, more likely to be charged more harshly, more likely to be punished, and, thus, more likely to enter into the criminal justice vortex where they're more likely to get into trouble in the future. One major problem is that Northpointe isn't actually assessing whether or not people are more likely to engage in criminal activity, but whether or not they are more likely to be arrested, charged, and convicted. They are relying on biased data and predicting outcomes that reinforced the biased system. And their predictions help create the outcomes that reinforce a biased system.

This is also the problem with predictive policing. We know from sociological work that whites are more likely to use AND sell drugs than blacks. Not just marijuana, but everything from coke to heroin. Yet, blacks are more likely to be arrested, charged, and convicted. And thus, when we feed arrest records back into the system, all signs point cops to go to poor black and brown neighborhoods to find criminal behavior. Predictive policing algorithms don't send cops to the university frat house because those people are not in the system. And because those people aren't in the system, they aren't presumed to be engaged in criminal behavior. And, thus, the system goes full loop and guarantees inequities continue.

In the criminal justice context, data is often used with actors knowing full well that they're prioritizing equality over equity. What many fail to realize is that they're not even achieving equality because they're lacking the data to achieve true equality. They don't know who is *not* in the system and violating the law; they're only making decisions based on who is there. And so bias is fed all the way through. And it's presumed to be better than the status quo, but, in effect, it's cementing the status quo. This is what happens when we simply focus on the available data and limit our purview to that narrow scope. We think we're doing good by making data available, but what we're doing is making available data that will continue structural divisions. Is that our goal?

The Cost of Feedback Loops

Many of you may be familiar with Latanya Sweeney's startling experiment, but if you're not, let me share it with you. As a computer scientist and the former

Chief Technologist of the FTC, Latanya has a good sense of how machine learning systems are designed and work. One day, she was doing an ego search on Google and she was served advertising for a criminal justice product. Curious if the algorithm targeted her in particular, she downloaded a list of popular baby names by race and ran a script to test if known black baby names received more criminal justice ads than known white baby names. They did. In poking around, she realized something important. Google doesn't sell ads based on the race of names, but it does evolve the targeting of its ads based on feedback loops. When people click on ads associated with a search term, the company tries to figure out what makes that term likely to work for a particular ad. All of this is done on the backend with no human-readable information. But because society is generally racist, people were more likely to click on criminal justice product ads when searching for black names. And so Google's system learned society's racism. It didn't need to know that it was categorizing names based on race or make any attempt to ask why. All Google needed was a matrix of correlations and it learned to spit back racist ideologies.

Categorization is fraught, especially when race is involved. If you haven't read it, I strongly recommend the book Sorting Things Out by Geof Bowker and Leigh Star, which highlights how racial categories during apartheid South Africa went terribly awry. Families split apart for having children darker than themselves. Before we pretend like we're better, let's keep in mind that anti-miscegenation laws in the US were based on the same logic.

Census has to deal with the challenges of racial categories every centennial. It's not easy to figure out how to do this right because it's all wrapped up in cultural logics. Worse, it's wrapped up in politics. The data that Census collects affects economic decisions (see: Native American communities) and shapes how politicians think about gerrymandering, not to mention the illegal practices of redlining that still go on. Census understands the political nature of their effort and works hard to develop solutions that get widespread buy-in. They don't just think the data is neutral; they know it's not. But the broader ecosystem isn't as mature in its thinking.

The problem with contemporary data analytics is that we're often categorizing people without providing human readable descriptors. Yes, the FTC caught some foolish data broker companies labeling segments of the population with titles like "Thrifty Elders" and "Urban Scramble", but most data analysis doesn't work that way. Most data analysis makes prejudicial decisions as part of clustering without having any understanding of the people or properties that they are using.

It is simply math. But that math - and the decisions that are determined by that math - have serious social ramifications.

If you want to get a job at Walmart, your resume will be filtered through a 3rd party applicant tracking system where it will be analyzed to see how your resume matches up against others who have succeeded at the job. Only those who are promising will be sent to the person in charge of hiring for consideration. The rest will be filtered. Although many of these systems do not explicitly judge people based on race or gender, plenty of markers in resumes are proxies for this. Gaps in employment, zipcode of address, etc. And most of the outcomes of these systems have a disparate impact. But unless you explicitly analyze for it, you probably don't know why. People get redlined without any form of redress.

There are interesting remedies for this. For example, a group of computer scientists have proposed a way to mathematically renormalize training data to minimize disparate impact. But this requires actually collecting sensitive data. And it requires wanting to achieve equity and combat bias. It's not clear that this is always what folks are aiming to do. The truth of the matter is that discriminatory hiring is actually more efficient. And if we're not careful, we'll allow technology to be used to enable such systems. As Cathy O'Neil argues, these are "weapons of math destruction."

Both this hiring case and the Google case highlight something important - transparency of an algorithm is not actually the solution. The problem is in the model, dependent on the training data and the evolution of the system in light of new data coming in. And when we let data systems learn from the public at large, when we allow feedback loops without thinking through the bias that emerges as a result, we allow data to be prejudicially shaped.

Towards Accountability

As we move towards open data and the use of more sophisticated algorithms, we need to start explicitly stating our values and grappling with accountability. Accountability isn't simple. In fact, one of the biggest problems right now is that we don't have the tools to do accountability well. Companies don't know how their systems will evolve based on user interaction. Google didn't design for black people to get criminal justice products. Facebook didn't design for conspiracy theorists to manipulate their algorithms. Walmart didn't hire a third party vendor to discriminate in employment on their behalf. But these large, well-funded companies don't even have the tools to know when they're being gamed, when their systems are being manipulated or used to do harm.

The government is in a different position than most corporations. If you get a ridiculous advertisement because of bad data, you'll laugh. I used to be labeled by many major systems as a trucker because of my fieldwork locations. The advertisements were priceless! But it's not so funny when you're sent to jail because a risk-assessment tool decided that you had a higher than average likelihood of re-offending because your father had been incarcerated. And it's not so funny when the schools in your community self-segregate and lead to increasing racial tensions that result in explosive riots.

Accountability takes work and thoughtfulness. Unfortunately, more often than not, we just look for someone to blame. That's not actually the same. Madeleine Elish was researching the history of autopilot in aviation when she uncovered intense debates about the role of a human pilot in autonomous systems. Not unlike what we hear today, there was tremendous pressure to keep pilots in the cockpit "in case of emergency." The idea was that, even as planes shifted from being primarily operated by pilots to primarily operated by computers, it was essential that pilots could step in last minute if something went wrong with the computer systems.

Although this was seen as a nod to human skill, what Madeleine saw unfold over time looked quite different. Pilots shifted from being skilled operators to being liability sponges. Time and time again, they were blamed when things went wrong and they failed to step in appropriately. Because they rarely flew, pilots' skills atrophied on the job, undermining their capabilities at a time when they became increasingly accountable. Because of this, Madeleine and a group of colleagues realized that the contexts in which humans are kept in the loop of autonomous systems can be described as "moral crumple zones," sites of liability in which the human is squashed when the socio-technical systems go wrong.

As we think about the importance of accountability in algorithmic systems, I want us to keep track of how certain decisions we make will have unexpected ripple effects. We need extensibility in our principles because we need to prepare for how solutions to current issues won't play out the way that we expect. Resistance and gaming will occur. Policies that seem to inform and educate will be deemed by future generations as bureaucratic overhead. Norms and standards of today will seem quaint tomorrow. We need to prepare for that.

Technology is increasingly becoming an arbitrator of social values. And as we build the tools for data, let's not lose track of that. We need to be attentive to the social factors and the dynamics of inequality that are shaping data analytics right

now. If we're not careful, we're more likely to build moral crumple zones than productive analytics systems.

I am excited about the possibility and future of using data to make wiser, more responsible decisions. Unfortunately, I don't have a lot of hope that this will be the driving goal when hype is dominating public rhetoric about the use of data. We have a responsibility to help the Commerce Department do right by their data and this means that we have a responsibility to make sure that they don't get too caught up in the hype. Commerce shouldn't be doing data work just to do data work. It should be doing so to make our country stronger. And, in my mind at least, I think we have a responsibility to make sure that our government uses data to combat inequities and prejudice along the way.

Thank you![1]

[1] D.S. Department of Commerce, Data Advisory Council: October 28, 2016

This talk was written for a meeting of the Data Advisory Council. It is a crib; the actual talk probably came out slightly differently.

*This study has been published with the approval of its writer.