AI Bias And The 'People Factor' In AI Development

forbes.com/sites/marksears1/2018/11/13/ai-bias-and-the-people-factor-in-ai-development/

Mark Sears
Contributor
AI & Big Data I write about AI, data and the future and science of work

Oscar Wilde once argued that life imitates art more than art imitates life. Strangely, that's proving to

In AI and machine learning programs, discrimination is caused by data. photo credit: Getty

Getty

be the case when it comes to AI development – but not in the way some had hoped.

On Star Trek: The Next Generation, the android Data was constantly endeavoring to evolve its programming to become more human. That's how AI works in our world now, as systems have advanced to the point where people are starting to envision what a workforce augmented by robots might look like. But at the same time AI has grown to become more like humans, a distinctly human roadblock has emerged in the application of the technology: bias.

Wasn't this supposed to be our shot to get it right? Since human bias doesn't appear to be going anywhere soon, technology was supposed to succeed at eliminating bias where human intelligence had failed miserably. Yet here we are, dealing with the same issues that we've dealt with in the humans-only world, bringing a crop of new challenges.

Addressing AI bias means understanding AI bias and, to do this, it's critical to understand how bias is introduced into AI systems.

An Algorithmic Impact

In AI and machine learning programs, discrimination is caused by data (not to be confused with <u>Data</u>). This "algorithmic bias" occurs when AI and computing systems act not in objective fairness, but according to the prejudices that exist with the people who formulated, cleaned and structured their data. This is not inherently harmful – human bias can be as simple as preferring red to blue – but warning signs have started to appear.

Earlier this year, a team of cross-disciplinary researchers at the <u>University of California Berkeley distinguished</u> biases pre-existing in training data from the technical biases that arise from the tools and algorithms that power these AI systems and from the emergent biases that result from human interactions with them.

AI is only as good as the data it is trained to analyze, which can include pre-existing (human) bias on the individual or societal level. This kind of <u>bias was found in a risk assessment software</u> known as COMPAS. Courtroom judges used it to forecast which criminals were most likely to offend. When news organization ProPublica compared COMPAS risk assessments for 10,000 people arrested in one county in Florida with data showing which ones went on to reoffend, it discovered that when the algorithm was right, its decision making was fair. But when the algorithm was wrong, people of color were almost twice as likely to be labeled a higher risk, yet they did not re-offend.

Technical bias arises from technical limitations, whether they are known or not. This can include the tools and algorithms an AI system uses. A May 2016 accident involving a Tesla Model S and a tractor trailer in Williston, Florida offers a strong example of technical bias. The accident killed the Tesla driver, who had autopilot engaged when a tractor trailer drove across a divided highway perpendicular to the car. Tesla later shared in a <u>blog article</u>, "Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied."

Emergent bias occurs only in the context of using the system, and enters into the picture when new knowledge is encountered or when there's a user and system design mismatch. The world saw emergent bias in action when a U.S. couple reported in May 2018 that one of

the Amazon Echo smart speakers in their home <u>recorded a private conversation and emailed it</u> to a friend without their knowledge. Amazon said the voice-activated Echo picked up a series of miscues during the conversation.

Naturally, these examples beg the question: What can be done to avoid AI bias? While it's too early to tell whether we can ever fully solve the problem, it's clear that the humans who play a critical role in creating AI systems will play a similarly critical one in addressing bias in those systems.

The Problem is the Solution

Unfortunately, we can't rely on technology to solve the equation of algorithm bias. No clever app is going to give AI systems the comprehension needed to spot and correct these errors. It's a people issue.

The position humans have in the AI stack is frequently misunderstood, as many aren't familiar with all that's happening behind the scenes. It takes an army of people spending countless hours creating algorithms and organizing extensive datasets to bring AI to life. People play a critical role in that process, so developers must build training and safeguards into their process to identify and reduce bias in AI systems. Since many would-be disruptors source at least some of their data annotation to third parties, this concept extends to their vendors and service providers.

Many resources are available to help developers with their data production lines these days, but not all tools are built equally. Some off-the-shelf crowdsourcing models, for example, come with inherit risks because they're <u>anonymous and unaccountable</u> to anyone. With no relationship established with the workers who process the data, there's no way to correct subtle problems that might emerge – leaving the doors open for bias to be introduced to important datasets. This renders those tools inadequate for any business looking to offload enterprise-grade work.

No matter how organizations decide to manage their data, due diligence is imperative to mitigate unintended bias. Accountability is a relationship-driven business model, so developers must identify ways to strategically deploy people in the data annotation process. Communication and the ability to evolve processes in real time are important for developers to ensure their AI systems consume training data that reflects accurate ground truth. They also must be able to initiate roadblocks and make improvements as necessary to eliminate potential bias in the data.

That Path Ahead

As long as people are developing AI technologies, bias is likely going to be a lingering issue. Similar to Data in Star Trek, the endeavor to reduce bias will be an evolution that requires us to strive for greatness. This means iterative development, constant testing and learning

about possible bias in developed systems, with accountability built into the process. Working closely with partners and building processes to identify and address bias will help all of us continue AI's advancement beyond what we previously thought possible – and possibly, even boldly going where no human has gone before.