

AI Fairness Isn't Just an Ethical Issue

by Greg Satell and Yassmin Abdel-Magied

October 20, 2020



Andriy Onufriyenko/Getty Images

Summary. There is often an assumption that technology is neutral, but the reality is far from it. Machine learning algorithms are created by people, who all have biases. They are never fully “objective”; rather they reflect the world view of those who build them. And unless there... [more](#)

The authority that administers A-Level college entrance exams in the UK, Ofqual, recently found itself mired in scandal. Unable to hold live exams because of Covid-19, it designed and employed an algorithm that based scores partly on the historical performance

of the schools students attended. The outcry was immediate, as students who were already disadvantaged found themselves further penalized by artificially deflated scores, their efforts disregarded and their futures thrown into disarray.

This is far from an isolated incident. Even the world's most sophisticated technology companies have faced similar problems. In 2018, Amazon's recruiting algorithm was flagged for penalizing applications that contained the word "women's." More recently, Apple's credit card algorithm was so biased against women that founder Steve Wozniak's wife was given a credit limit 10 times lower than his, despite them sharing all assets and accounts.

INSIGHT CENTER

AI and Equality

Designing systems that are fair for all.

Clearly, this is unfair. Just as important, these biases undermine efficiency and productivity. If we see a value in putting the best students in the best schools, the inherent bias in

Ofqual's algorithm undermines that purpose. Similarly, a biased recruiting algorithm undermines a firm's ability to attract the best talent, and a biased credit rating algorithm undermines the ability to make smart credit decisions.

Just as the costs of bias are substantial, the benefits of eliminating bias can be just as significant. In fact, one econometric study at Stanford University found that at least "25% of growth in U.S. GDP between 1960 and 2010 can be attributed to greater gender and racial balance in the workplace," and that the figure could be as high as 40%.

Sources of Bias

In order to fully unleash the profitable opportunities of AI, we first need to understand where biases come from. There is often an assumption that technology is neutral, but the reality is far

from it. Machine learning algorithms are created by people, who all have biases. They are never fully “objective”; rather they reflect the world view of those who build them and the data they’re fed.

Biased human judgments can affect AI systems in two different ways. The first is bias in the data that systems learn from. You can see this play out for yourself: do a Google image search for “professional haircut,” and another for “unprofessional haircut.” “Professional haircut” turns up results that are exclusively white men, while “unprofessional haircut” has much more gender and racial diversity. (This issue was originally surfaced by Twitter users back in 2016.)

Is it really true that only white men have professional haircuts? Of course not. The Google results are based on articles written about professional haircuts. They reflect human editorial decisions to include and prioritize white men. So a supposedly “neutral” search provides a decidedly unneutral outcome. Bias based on historical norms is fairly common.

A second source of bias occurs in the way algorithms are designed. Machine learning algorithms generally rely on correlation, rather than causal relationships. For example, it’s relatively easy to design an algorithm that will find a relationship between temperature readings and ice cream consumption, showing that as thermometer readings go up, people eat more ice cream.

Of course, high temperature readings themselves don’t actually *cause* increased ice cream sales. Putting a thermometer next to a heater will not elicit a sudden rush for something cold. Humans get that. Algorithms often do not. The problem is that there is an intermediate factor, called a confounding factor, that affects both thermometer readings and ice cream sales, namely hot weather, that algorithms are not typically designed to take into account.

Ice cream is a fairly trivial example. But what if correlations between a criminal offender's zip code and recidivism rates, with poverty as a confounding factor, were used to determine prison terms? Would we be sentencing people to longer prison sentences merely because they were poor? Sadly, there is considerable evidence that this injustice is happening right now, and not just with recidivism rates.

Researchers, led by artificial intelligence pioneer Judea Pearl, are working to mitigate this type of bias through developing causal AI techniques, but for now at least, our systems are rife with algorithmic bias.

Mitigating Systemic AI Bias

Unless there is concerted intervention, algorithms will continue to reflect and reinforce the prejudices that hold society and business back. For example, developers should no longer be able to simply insert open source code into systems without knowing what's in it, and software firms should no longer be able to use the concept of "proprietary algorithms" to obfuscate scrutiny of shoddy code design.

We can preempt some of the damage by utilizing ethical AI design principles. We also need to ensure that our algorithms are explainable, auditable, and transparent. Just as we wouldn't accept humans making major decisions that affect others without any oversight or accountability, we should not accept it from algorithms.

Here, the Ofqual case is instructive. Because the agency was transparent about how the algorithm was constructed, the source of the bias was quickly revealed, corrective action was taken in a timely manner, and much of the damage was likely mitigated. As Linus's Law states, "given enough eyeballs, all bugs are shallow."

Every AI system needs to be held accountable. We can't lose sight of the fact that AI systems are tools. They need to be designed to benefit humans, not harm them.

Unlocking Value by Mitigating Bias

One example of the enormous effect systemic bias can have is the market for credit. In the U.S. particularly, there is an ugly history of redlining that discriminated against racial minorities and, to a large extent, effectively locked unfavored groups out of the financial system.

However, banks don't earn money by turning away performing loans; they earn money by extending loans to good credit risks. It was this simple principle that led the credit bureau Experian to develop its Boost program, which allows so-called "thin file" clients with limited credit history to augment their scores with information about their history paying things like cable and mobile phone bills.

Since the program launched in January 2019, more than 4 million people have participated. Sixty-one percent of those with "poor" credit — more than 150,000 people — have been able to upgrade to a "fair" rating, giving them far more access to buying a car, renting or buying a home, or acquiring a business loan. Lenders are also able to extend significantly more credit to profitable customers.

The company hopes to build on that success with a model that mitigates algorithmic and data biases which affect the evaluation of credit risks. "We believe that by providing fair and equitable assessment for consumers, we are also helping lenders make more informed and profitable decisions," says Eric Haller, Head of Experian Datalabs. "Everybody wins."

Opportunities from minimizing bias extend far beyond the credit industry. We need to start looking at eliminating AI bias less as merely a "nice thing to do," and more as an economic and

competitive imperative. Business leaders take note: By making our AI systems more fair, we can also make our organizations more profitable and productive.

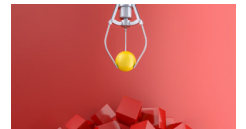
Greg Satell is an international keynote speaker, adviser and bestselling author of *Cascades: How to Create a Movement that Drives Transformational Change*. His previous effort, *Mapping Innovation*, was selected as one of the best business books of 2017. You can learn more about Greg on his website, GregSatell.com and follow him on Twitter @DigitalTonto.

YA

Yassmin Abdel-Magied is a writer and broadcaster. A LinkedIn Changemaker, Yassmin is a globally sought-after advisor on unconscious bias, inclusive leadership, and social justice. Yassmin's TED talk, *What does my headscarf mean to you*, was chosen as one of TED's top ten ideas of 2015. Learn more on her website or follow her on Twitter.

Recommended For You

What Do We Do About the Biases in AI?



We Need AI That Is Explainable, Auditable, and Transparent



What Does Building a Fair AI Really Entail?





How to Fight Discrimination in AI

