Eliminating bias is the new frontier of AI innovation

By Rudeon Snell   9 Sep 2020

2020 is forcing us to confront some hard truths about the world we live in. The Covid-19 pandemic has cast a sobering spotlight on the unsustainable path we are on.

Rudeon Snell, Global Senior Director: Industries & Customer Advisory at SAP

One such truth is symbolised by the global #BlackLivesMatter movement, which has once again highlighted the embedded biases in our interconnected social fabric, forcing us all, to re-evaluate long standing notions of morality, fairness and ethics.

It is worth taking pause, to consider whether the exponential technological progress is not also amplifying some of the very same challenges we are trying to overcome, as a global society.

As we strive to meet the unmet and unarticulated needs of customers, we continuously look towards technology to help meet those needs. We see leading companies globally investing heavily in technologies such as cloud computing, internet of things, advanced analytics, edge computing, virtual and augmented reality, 3D printing and of course artificial intelligence. And it is AI, that many experts tout as one of the most transformational technologies of our time, in terms of sheer transformational impact on humanity.

Global use of AI has ballooned by 270% over the past five years, with estimated revenues of more than $118bn by 2025. AI powered technology solutions have become so pervasive, a recent Gallup poll found that nearly 9 in 10 Americans use AI based solutions in their everyday lives.

And yet, a dark side of AI is surfacing with alarming frequency as AI engrains itself in our daily lives.

The dark side of AI

In 2018, reports emerged of Gmail's predictive text capability automatically assigning "investor" as "male". When a research scientist typed "I am meeting an investor next week", Gmail's Smart Compose tool thought they would want to follow up with the question: "Do you want to meet him?"

That same year, Amazon had to decommission its AI-powered talent acquisition system after it appeared to favour male candidates. The software seemingly downgraded female candidates if their resumes included phrases with the word "women's" in them, for example "women's hockey club captain."

Errant algorithms can cause greater harm than a few missed employment opportunities.

In June 2020, the *New York Times* reported on an African American man wrongfully arrested for a crime he didn't commit after a flawed match from a facial recognition algorithm.

Recent studies by MIT found that facial recognition software, used by US police departments for decades, work relatively well on certain demographics, but is far less effective on other demographics, mainly due to a lack of diversity in the data that the developers used to train these algorithms.

Microsoft and Amazon have halted sales of their facial recognition software until there is a better understanding and mitigation of their impact, on especially vulnerable or minority communities. IBM has even gone as far to halt offering, developing or researching facial recognition technology.

What causes bias in AI?

There is growing evidence that it is the underlying data that perpetuates bias in AI. For example, using news articles for natural language processing could instil the common gender stereotypes we find in society simply due to the nature of the language used.

Many of the early algorithms were also trained using web data, which is often rife with our raw, unfiltered thoughts and prejudices. A person commenting anonymously on an online forum arguably has more freedom to display prejudices without much consequence. Any algorithm trained on this data is likely to assimilate the embedded biases.

As Princeton researcher Olga Russakovsky says: "Debiasing humans is a lot harder than debiasing AI systems."

Steps to greater fairness

There is arguably a need for greater diversity in the development rooms where AI algorithms are created. A cursory glance at the demographics of the big tech firms shows a disproportionate gender and demographic bias. More must be done to accelerate the synthesis of diverse and inclusive perspectives in the AI creation process, so that AI algorithms and the data they are trained on embody a broad range of perspectives, allowing them to drive more optimal outcomes for all those represented in society.

What can we do to mitigate bias in the AI solutions we increasingly use to make potentially life changing decisions? Greater

awareness of bias can help developers see the context in which AI could amplify embedded bias and guide them to put corrective measures in place. Testing processes should also be developed with bias in mind: AI creators should deliberately create processes and practices that test for and correct bias.

Finally, AI firms need to make investments into bias research, partnering with other disciplines far beyond technology such as psychology or philosophy, and share the learnings broadly to ensure all the algorithms we use can operate alongside humans in a responsible and helpful manner.

Fixing bias is not something we can do overnight. It's a process, just like solving discrimination in any other part of society. However, with greater awareness and a purposeful approach to combating bias, AI algorithm creators have a hugely influential role to play in helping establish a more fair and just society for everyone.

This could be one silver lining in the ominous cloud that is 2020.

ABOUT RUDEON SNELL

Director: Intelligent Solutions at SAP EMEA South Africa
▪ Eliminating bias is the new frontier of AI innovation - 9 Sep 2020
▪ How AI-powered chatbots are unlocking business value - 14 Dec 2018
▪ Ethical considerations for AI - 5 Sep 2018

View my profile and articles...

For more, visit: https://www.bizcommunity.com