

IBM: If AI decision needs closer look, stay woke, and here's how

21 September 2018, by Nancy Owano



Credit: CC0 Public Domain

OK we get it. The gee-whiz feeling of artificial intelligence as shown in its developmental glory is with us and we ride the wave willingly.

A sobering phase though is definitely in the wings. Now that we have AI, what are we doing with it, and are we managing, even assessing, it well?

"Like video gamers looking for the next hack, employees will need to monitor, understand, question and exploit the vulnerabilities of their tools and account for them," said John Sumser, principal analyst at *HR Examiner*.

"Digital employees are central to our future, but managing them is very different than managing people or older software." Quoted in *Human Resource Executive*: "...understand that we are at the beginning of building and [using](#) intelligent tools, there is much work ahead and we will have to think about our machines differently from now on."

The AI supporters for decisions as used by governments and large organizations, after all, do impact our lives.

The big question is, who and what are training AI to make decisions? Is there [bias](#) baked into the training phase? If so, how can be sure the result is the fairest [decision](#)?

Long and short, IBM researchers have been busy devising ways to reduce bias in the datasets used to train AI. What are they doing? Are we to just look at yet another white paper? They are doing more than that.

They are delivering a rating system that can rank the relative fairness of an AI system.

Fairness is not just something that caught the attention of IBM. Zoe Kleinman, technology reporter, BBC News, wrote, "There is increasing concern that algorithms used by both tech giants and other [firms](#) are not always fair in their [decision-making](#)."

IBM's arsenal of [artificial intelligence](#) technology tools now includes a way to ferret out unconscious bias in making decisions. Bias doesn't always come dressed in neon lights and magic-marker labels. Half the time we are even cross-examining our own ability to judge, feeling uneasy about the other half of us that suspects the decision was rigged with bias. Make no mistake, though, our sniffers are often correct.

"Any number of predispositions can be baked into an [algorithm](#), hidden in a data set or somehow conceived during a [project](#)'s execution," said Jack Murtha on Wednesday in *Healthcare Analytics News*.

IBM is making this week's AI-related news.

IBM has announced a software service running on IBM Cloud that can detect bias and explains how AI makes decisions—as the decisions are being made, Murtha said.

"We are giving new transparency and control to the businesses who use AI and face the most potential risk from any flawed decision making," the general manager of Watson AI at IBM, Beth Smith, stated.

"Customers will be able to see, via a visual dashboard, how their algorithms are making decisions and which factors are being used in making the final recommendations," said Kleinman.

The IBM cloud-based software will be open-source, and will work with some commonly used frameworks for building algorithms. So what will it actually do?

Murtha fleshed it out. (1) It pings "unfair outcomes" realtime and (2) recommends data that could mitigate bias; (3) IBM is also offering consulting services to scrub decision making via stronger business processes and human-AI interfaces.

IBM's new contribution could add yet another layer to understanding and addressing bias.

Unfairness may be reflected in lack of diversity going into amounts of data that algorithms are trained on.

A CNBC report noted that "The composition of the tech industry that is creating those algorithms was not perfect. "Silicon Valley has a long history of being criticized for its lack of diversity."

Kay Firth-Butterfield, head of AI and machine learning at the World Economic Forum, was quoted by CNBC.

"When we're talking about bias, we're worrying first of all about the focus of the people who are creating the algorithms," Firth-Butterfield said. "We need to make the industry much more diverse in the [West](#)."

A postgraduate student at Massachusetts Institute of Technology in 2016 had found that "facial recognition only spotted her [face](#) if she wore a white mask," said Kleinman.

What's next? "IBM Services will work with [businesses](#) to help them utilize the new service.

IBM Research will release a toolkit into the open source community," said *Seeking Alpha*. *ZDNet* had more details to share about this [toolkit](#). IBM will open source "bias detection tools" from IBM Research via an "AI Fairness 360 toolkit." Expect to see a library of algorithms, code and tutorials.

ZDNet's Larry Dignan: "The hope is that academics, researchers and data scientists will integrate bias detection into their models."

Those who want to delve more into such a toolkit can check out IBM's tools on [Github](#).

© 2018 Tech Xplore

APA citation: IBM: If AI decision needs closer look, stay woke, and here's how (2018, September 21)
retrieved 25 January 2022 from <https://techxplore.com/news/2018-09-ibm-ai-decision-closer-woke.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.