Microsoft is creating an oracle for catching biased AI algorithms

Notes & Cues:

Article:

Microsoft is building a tool to automatically identify bias in a range of different AI algorithms. It is the boldest effort yet to automate the detection of unfairness that may creep into machine learning—and it could help businesses make use of AI without inadvertently discriminating against certain people.

Algorithmic bias is a growing concern for many researchers and technology experts. As algorithms are used to automate important decisions, there is a risk that bias could become automated, deployed at scale, and more difficult for the victims to spot.

"Things like transparency, intelligibility, and explanation are new enough to the field that few of us have sufficient experience to know everything we should look for and all the ways that bias might lurk in our models, " says Rich Caruna, a senior researcher at Microsoft who is working on the bias-detection dashboard.

Facebook announced its own tool for detecting bias at its annual developer conference on May 2. Its tool, called Fairness Flow, automatically warns if an algorithm is making an unfair judgement about someone based on his or her race, gender, or age.

Bin Yu, a professor at UC Berkeley, says the tools from Facebook and Microsoft seem like a step in the right direction, but may not be enough. She suggests that big companies should have outside experts audit their algorithms in order to prove they are not biased. "Someone else has to investigate Facebook's algorithms—they can't be a secret to everyone, " Yu says.

Summary: