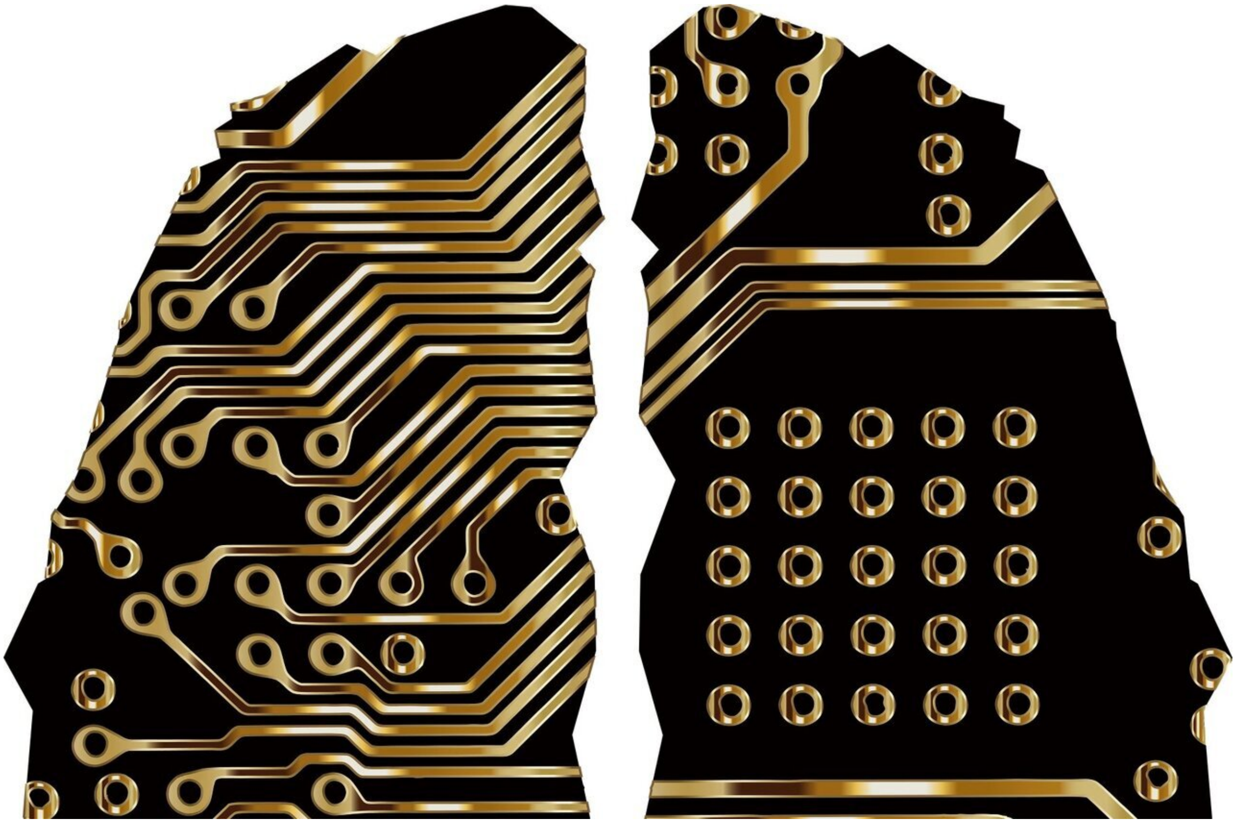
TechXplore

Machine learning has a flaw; it's gullible

June 23 2020



Credit: Pixabay/CC0 Public Domain

Artificial intelligence and machine learning technologies are poised to supercharge productivity in the knowledge economy, transforming the future of work.

But they're far from perfect.

Machine learning (ML)—technology in which algorithms "learn" from existing patterns in data to conduct statistically driven predictions and facilitate decisions—has been found in multiple contexts to reveal bias. Remember when Amazon.com came under fire for a hiring algorithm that revealed gender and racial bias? Such biases often result from slanted training data or skewed algorithms.

And in other business contexts, there's another potential source of bias. It comes when outside individuals stand to benefit from bias predictions, and work to strategically alter the inputs. In other words, they're gaming the ML systems.

It happens. A couple of the most common contexts are perhaps job applicants and people making a claim against their insurance.

ML algorithms are built for these contexts. They can review resumes way faster than any recruiter can, and can comb through insurance claims faster than any human processor.

But people who submit resumes and insurance claims have a strategic interest in getting positive outcomes—and some of them know how to outthink the algorithm.

This had researchers at the University of Maryland's Robert H. Smith School of Business wondering, "Can ML correct for such strategic behavior?"

In new research, Maryland Smith's Rajshree Agarwal and Evan Starr, along with Harvard's Prithwiraj Choudhury, explore the potential biases that limit the effectiveness of ML process technologies and the scope for human capital to be complementary in reducing such biases. Prior research in so-called "adversarial" ML looked closely at attempts to "trick" ML technologies, and generally concluded that it's extremely challenging to prepare the ML technology to account for every possible input and manipulation. In other words, ML is trickable.

What should firms do about it? Can they limit ML prediction bias? And, is there a role for humans to work with ML to do so?

Starr, Agarwal and Choudhury honed their focus on patent examination, a context rife with potential trickery.

"Patent examiners face a time-consuming challenge of accurately determining the novelty and nonobviousness of a patent application by sifting through ever-expanding amounts of 'prior art,'" or inventions that have come before, the researchers explain. It's challenging work.

Compounding the challenge: patent applicants are permitted by law to create hyphenated words and assign new meaning to existing words to describe their inventions. It's an opportunity, the researchers explain, for applicants to strategically write their applications in a strategic, ML-targeting way.

The U.S. Patent and Trademark Office is generally wise to this. It has invited in ML technology that "reads" the text of applications, with the goal of spotting the most relevant prior art quicker and leading to more accurate decisions. "Although it is theoretically feasible for ML algorithms to continually learn and correct for ways that patent applicants attempt to manipulate the algorithm, the potential for patent applicants to dynamically update their writing strategies makes it practically impossible to adversarially train an ML algorithm to correct for this behavior," the researchers write.

In its study, the team conducted observational and experimental research. They found that patent language changes over time, making it highly challenging for any ML tool to operate perfectly on its own. The

ML benefitted strongly, they found, from human collaboration.

People with skills and knowledge accumulated through prior learning within a domain complement ML in mitigating bias stemming from applicant manipulation, the researchers found, because domain experts bring relevant outside information to correct for strategically altered inputs. And individuals with vintage-specific skills—skills and knowledge accumulated through prior familiarity of tasks with the technology—are better able to handle the complexities in ML technology interfaces.

They caution that although the provision of expert advice and vintage-specific human capital increases initial productivity, it remains unclear whether constant exposure and learning-by-doing by workers would cause the relative differences between the groups to grow or shrink over time. They encourage further research into the evolution in the productivity of all ML technologies, and their contingencies.

More information: Prithwiraj Choudhury et al, Machine learning and human capital complementarities: Experimental evidence on bias mitigation, *Strategic Management Journal* (2020). DOI: 10.1002/smj.3152

Provided by University of Maryland

Citation: Machine learning has a flaw; it's gullible (2020, June 23) retrieved 27 April 2024 from https://techxplore.com/news/2020-06-machine-flaw-gullible.html