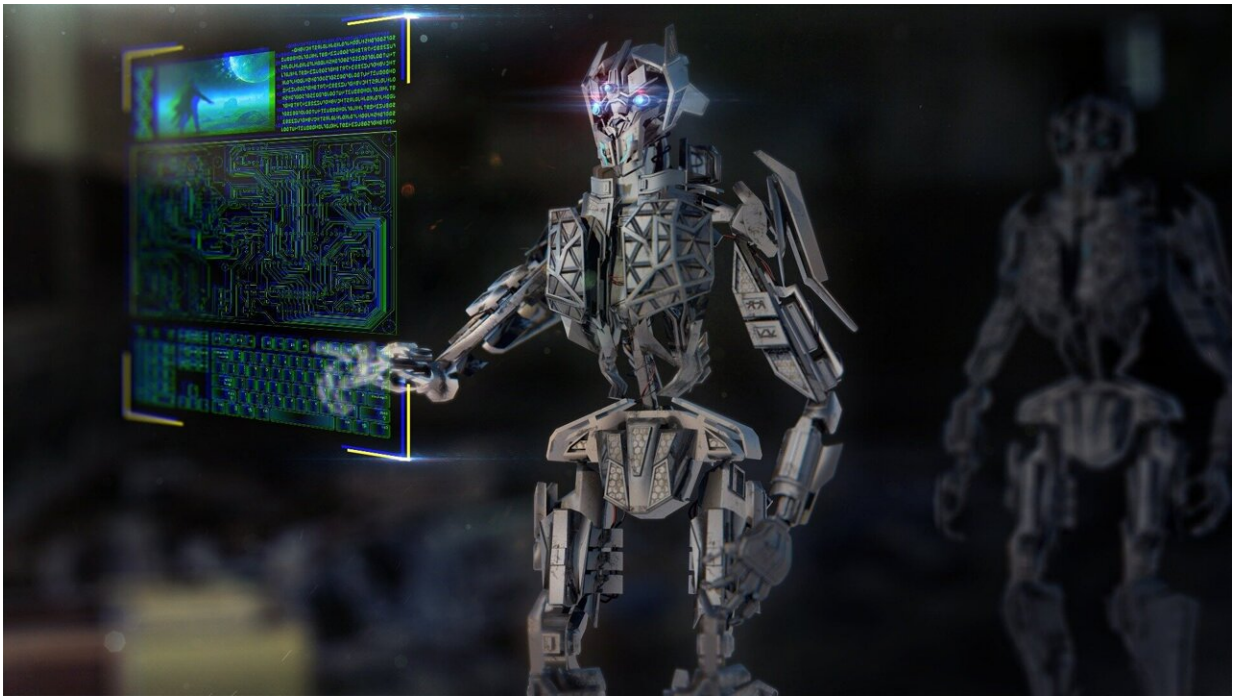
Study sheds light on the dark side of AI

April 6 2023, by Geoff McMaster



Credit: Pixabay/CC0 Public Domain

Artificial intelligence is touted as a panacea for almost every computational problem these days, from medical diagnostics to driverless cars to fraud prevention.

But when AI fails, it does so "quite spectacularly," says Vern Glaser of the Alberta School of Business. In his recent study, "When Algorithms Rule, Values Can Wither," Glaser explains how human values are often

subsumed by AI's efficiency imperative, and why the costs can be high.

"If you don't actively try to think through the value implications, it's going to end up creating bad outcomes," he says.

When bots go bad

Glaser cites Microsoft's Tay as one example of bad outcomes. When the chatbot was introduced on Twitter in 2016, it was revoked within 24 hours after trolls taught it to spew racist language.

Then there was the "robodebt" scandal of 2015, when the Australian government used AI to identify overpayments of unemployment and disability benefits. But the algorithm presumed every discrepancy reflected an overpayment and automatically sent notification letters demanding repayment. If someone didn't respond, the case was forwarded to a debt collector.

By 2019, the program identified more than 734,000 overpayments worth two billion Australian dollars (C$1.8 billion).

"The idea was that by eliminating human judgment, which is shaped by biases and personal values, the automated program would make better, fairer and more rational decisions at much lower cost," says Glaser.

But the human consequences were dire, he says. Parliamentary reviews found "a fundamental lack of procedural fairness" and called the program "incredibly disempowering to those people who had been affected, causing significant emotional trauma, stress and shame," including at least two suicides.

While AI promises to bring enormous benefits to society, we are now also beginning to see its dark underbelly, says Glaser. In a recent Globe

and Mail column, Lawrence Martin points out AI's dystopian possibilities, including autonomous weapons that can fire without human supervision, cyberattacks, deepfakes and disinformation campaigns. Former Google CEO Eric Schmidt has warned that AI could quite easily be used to construct killer biological weapons.

Glaser roots his analysis in French philosopher Jacques Ellul's notion of "technique," offered in his 1954 book "The Technological Society," by which every field of human activity is determined by the imperatives of efficiency and productivity.

"Ellul was very prescient," says Glaser. "His argument is that when you're going through this process of technique, you are inherently stripping away values and creating this mechanistic world where your values essentially get reduced to efficiency.

"It doesn't matter whether it's AI or not. AI in many ways is perhaps only the ultimate example of it."

A principled approach to AI

To guard against the "tyranny of technique" in AI, Glaser suggests adherence to three principles. First, recognize that because algorithms are mathematical, they rely on "proxies," or digital representations of real phenomena.

One way Facebook gauges friendship, for example, is by how many friends a user has, or by the number of likes received on posts from friends.

"Is that really a measure of friendship? It's a measure of something, but whether it's actually friendship is another matter," says Glaser, adding that the intensity, nature, nuance and complexity of human relationships

can easily be overlooked.

"When you're digitizing phenomena, you're essentially representing something as a number. And when you get this kind of operationalization, it's easy to forget it's a stripped-down version of whatever the broader concept is."

For AI designers, Glaser recommends strategically inserting human interventions into algorithmic decision-making, and creating evaluative systems that account for multiple values.

"There's a tendency when people implement algorithmic decision-making to do it once and then let it go," he says, but AI that embodies human values requires vigilant and continuous oversight to prevent its ugly potential from emerging.

In other words, AI is simply a reflection of who we are—at our best and our worst. Without a good, hard look in the mirror, the latter could take over.

"We want to make sure we understand what's going on, so the AI doesn't manage us," he says. "It's important to keep the dark side in mind. If we can do that, it can be a force for social good."

More information: When Algorithms Rule, Values Can Wither. sloanreview.mit.edu/article/wh … e-values-can-wither/

Provided by University of Alberta

Citation: Study sheds light on the dark side of AI (2023, April 6) retrieved 16 August 2024 from https://techxplore.com/news/2023-04-dark-side-ai.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.