

Study takes aim at biased AI facial-recognition technology

February 12 2019, by Liz Do



A recent study by Deb Raji and researchers at the MIT Media Lab shows a need for stronger evaluation practices of AI products to mitigate gender and racial biases. Credit: Liz Do

A study by Deb Raji, a fourth-year student in the University of Toronto's

Faculty of Applied Science & Engineering, and researchers at the Massachusetts Institute of Technology is underscoring the racial and gender biases found in facial-recognition services.

Raji spent the summer of 2018 as an intern at MIT's Media Lab, where she audited commercial facial-recognition technologies made by leading companies such as Microsoft, IBM and Amazon. The researchers discovered that all of them had a tendency to mistake darker-skinned women for men.

But one service in particular – Amazon's Rekognition – showed a higher level of bias than the rest. Although it could identify the gender of light-skinned men with nearly 100 per cent accuracy, it misclassified women as men 29 per cent of the time, and darker-skinned women for men 31 per cent of the time.

Rekognition was recently piloted by police in Orlando, Fla., using the service in policing scenarios such as scanning faces on cameras and matching them against those in criminal databases.

"The fact that the technology doesn't characterize Black faces well could lead to misidentification of suspects," says Raji. "Amazon is due for some public pressure, given the high-stakes scenarios in which they're using this technology."

With rapid advancements and deployment of artificial intelligence (AI) products, this new study emphasizes the need to not only test systems for performance, but also for potential biases against underrepresented groups.

Although algorithms should be neutral, Raji explains that because [data sets](#) – information used to "train" an AI model – are sourced from a society that still grapples with everyday biases, these biases become

embedded into the algorithms.

"Let's say I want examples of what healthy skin looks like. If you Google it now, you will see mostly light-skinned women," says Raji. "You won't see a man for pages, and you wouldn't see a darker-skinned woman until you really scroll down. If you feed that into an AI model, it adopts this world view and adapts its decisions based on those biases."

These biases should be called out, just as one would hold a person accountable, says Raji. "There's this increased danger when you embed that bias into an algorithm versus when a human makes a prejudiced decision. Someone will tell you it's wrong, whether it's the public or your boss," she says.

"With AI, we tend to absolve this responsibility. No one is going to put an algorithm in jail."

Raji's passion on the subject of [bias](#) in [machine learning](#) comes from her time as a Professional Experience Year Co-op (PEY Co-op) student at the AI startup Clarifai, where the topic of AI and ethics was regularly discussed at the research-oriented company.

"It's something that the company noticed and was very explicit about addressing, and it's a subject that personally resonated with me because I'm a visible minority," she says.

It also stems from her very own personal experiences with racially biased technologies. "I'd build something at a hackathon and wonder why it couldn't detect my face, or why an automated faucet can't detect my hand," she says.

Raji shared her experiences with computer scientist and digital activist, Joy Buolamwini, at MIT's Media Lab. This led to the internship, and to

Raji becoming the lead author on a paper that she presented at the Association for the Advancement of Artificial Intelligence Conference on AI Ethics and Society.

"I know it looks like I wrote a research paper in three months," says Raji. "But this issue has been percolating inside of me for much longer."

Raji is currently finishing her last term in [engineering science](#) and running a student-led initiative called Project Include, which trains students to teach computer programming in low income neighbourhoods in Toronto and Mississauga. She is also a mentee at Google AI. As part the mentorship program, she is working on a new thesis that focuses on practical solutions to hold companies accountable.

"People sometimes downplay the urgency by saying, 'Well, AI is just so new,'" says Raji. "But if you're building a bridge, would the industry allow you to cut corners and make those kinds of excuses?"

More information: Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. www.aies-conference.com/wp-con ... IES-19_paper_223.pdf

Provided by University of Toronto

Citation: Study takes aim at biased AI facial-recognition technology (2019, February 12) retrieved 24 April 2024 from <https://phys.org/news/2019-02-aim-biased-ai-facial-recognition-technology.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.