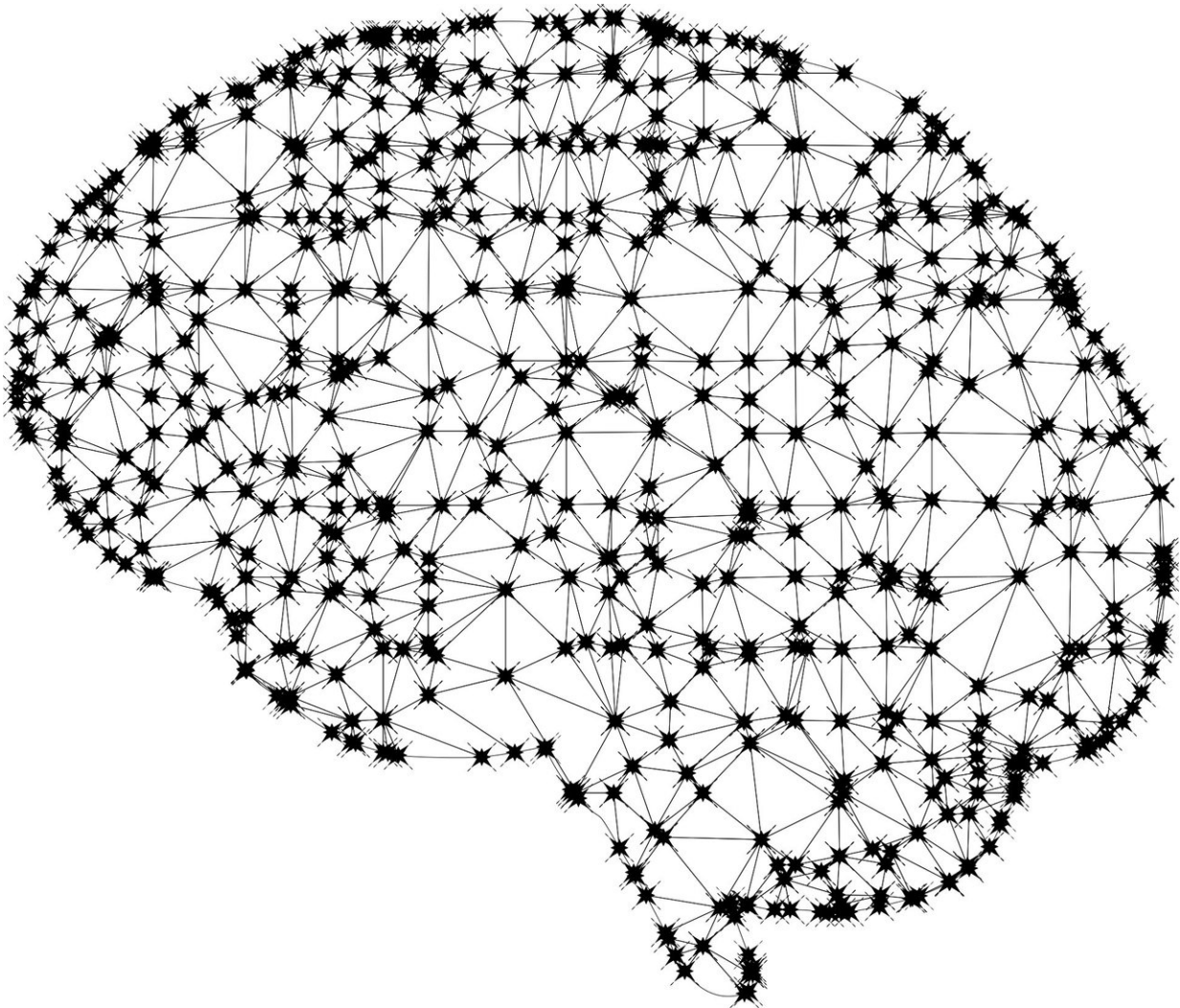


When bias in applicant screening AI is necessary

May 18 2020, by Sherry Stokes



Credit: CC0 Public Domain

Some biases in AI might be necessary to satisfy critical business requirements, but how do we know if an AI recommendation is biased strictly for business necessities and not other reasons?

A company receives 1000 applications for a new position, but whom should it hire? How likely is a criminal to become a repeat offender if they are released from prison early? As [artificial intelligence](#) (AI) increasingly enters our lives, it can help answer those questions. But how can we manage the biases that are in the [data sets](#) that AI uses?

"AI decisions are tailored to the data that is available around us, and there have always been biases in data, with regards to race, gender, nationality, and other protected attributes. When AI makes decisions, it inherently acquires or reinforces those biases," says Sanghamitra Dutta, a doctoral candidate in electrical and computer engineering (ECE) at Carnegie Mellon University.

"For instance, zip codes have been found to propagate racial [bias](#). Similarly, an automated hiring tool might learn to downgrade women's resumes if they contain phrases like "women's rugby team," say Dutta. To address this, a large body of research has developed in the past decade that focuses on fairness in machine learning and removing bias from AI models.

"However, some biases in AI might need to be exempted to satisfy critical business requirements," says Pulkit Grover, a professor in ECE who is working with Dutta to understand how to apply AI to fairly screen job applicants, among other applications.

"At first, it may seem strange, even politically incorrect, to say that some biases are okay, but there are situations where common sense dictates that allowing some bias might be acceptable. For instance, firefighters need to lift victims and carry them out of burning buildings. The ability

to lift weight is a critical job requirement," says Grover.

In this example, the capacity to lift heavy weight may be biased towards men. "This is an example where you may have bias, but it is explainable by a safety-critical, business necessity," says Grover.

"The question then becomes how do you check if an AI tool is giving a recommendation that is biased purely due to business necessities and not other reasons." Alternatively, how do you generate new AI algorithms whose recommendations are biased only due to business necessity? These are important questions relevant to U.S. laws on [employment discrimination](#). If an employer can show that a feature, such as the need to lift bodies, is a bona fide occupational qualification, then that bias is exempted by law. (This is known as "Title VII's business necessity defense.")

AI algorithms have become amazingly good at identifying patterns in the data. This ability, if left unchecked, can lead to unfairness due to stereotyping. AI tools, therefore, must be able to explain and defend the recommendations they are making. The team used their novel measure to train AI models to weed through biased data and remove biases that are not critical to perform a job while leaving those biases considered business necessary.

According to Dutta, there are some technical challenges in using their measure and models, but those can be overcome, as the team has demonstrated. However, there are important social questions to address. One key point is that their model can't automatically determine which features are [business](#) critical. "Defining the critical features for a particular application is not a mere math problem which is why [computer scientists](#) and [social scientists](#) need to collaborate to expand the role of AI in ethical employment practices," Dutta explained.

In addition to Dutta and Grover, the research team consists of Anupam Datta, professor of ECE; Piotr Mardziel, systems scientist in ECE; and Ph.D. candidate Praveen Venkatesh.

Dutta presented their research in a paper called, "An Information-Theoretic Quantification of Discrimination with Exempt Features," at the 2020 AAAI Conference on Artificial Intelligence in New York City.

More information: An Information-Theoretic Quantification of Discrimination with Exempt Features, Sanghamitra Dutta, Praveen Venkatesh, Piotr Mardziel, Anupam Datta, Pulkrit Grover, aaai.org/Papers/AAAI/2020GB/AAAI-DuttaS.9451.pdf

Provided by Carnegie Mellon University

Citation: When bias in applicant screening AI is necessary (2020, May 18) retrieved 16 April 2024 from <https://techxplore.com/news/2020-05-bias-applicant-screening-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.