

Algorithms that adjust for worker race, gender still show biases





Credit: *Data Mining and Knowledge Discovery* (2023). DOI: 10.1007/s10618-022-00910-8

Even after algorithms are adjusted for overt hiring discrimination, they may show a subtler kind: preferring workers who mirror dominant groups, according to a new study from researchers at The University of



Texas at Austin.

For example, when recruiting in a field that has more men, algorithms may favor people who more closely resemble masculine stereotypes. This tendency, or "social norm bias," compounds existing patterns in the workplace.

There may be a better approach for employers looking for technological fixes to combat algorithm bias—the tendency for hiring and recruiting algorithms to screen out job applicants by race or gender, according to study author Maria De-Arteaga, an assistant professor of information, risk and <u>operations management</u> in the McCombs School of Business.

The research is online in advance in *Data Mining and Knowledge Discovery*.

With Myra Cheng of Stanford University and Adam Tauman Kalai and Lester Mackey of Microsoft Research, De-Arteaga tested three common techniques for making algorithms fairer. The researchers used a data set with 397,340 biographies, spanning 28 occupations. Because the biographies were in third person, each biography had a "she" or "he" pronoun associated with it. An additional data set of biographies used nonbinary pronouns.

They then applied the three different types of interventions. The central question: Would the adjusted results display social norm bias when using someone's biography to predict the person's occupation?

Unfortunately, the bias persisted. For male-dominated occupations, the algorithm looked for language associated with men. If it didn't find such language, it was less accurate in guessing a person's occupation.

For example, the algorithm associated the word "empowerment" with



women. Female surgeons who used the word in their biographies were less likely to be identified as surgeons.

"When there is social norm bias, the individuals in the minority who benefit from an intervention will be those who most adhere to the social norms of the majority," De-Arteaga said. "The findings have widespread implications for correcting algorithm bias."

Using current techniques, companies may think they have addressed gender discrimination. But because those techniques are based on rigid characteristics associated with a group, they don't show the whole picture. They may penalize people who don't fit stereotypes of the majority.

To help compensate for those problems, De-Arteaga and her colleagues propose a formula to directly measure the degree of social norm <u>bias</u> in an algorithm. Data science or machine learning departments could use the formula to guide <u>algorithm</u> selection, the researchers said.

More information: Myra Cheng et al, Social norm bias: residual harms of fairness-aware algorithms, *Data Mining and Knowledge Discovery* (2023). DOI: 10.1007/s10618-022-00910-8

Provided by University of Texas at Austin

Citation: Algorithms that adjust for worker race, gender still show biases (2023, February 8) retrieved 7 May 2024 from <u>https://techxplore.com/news/2023-02-algorithms-adjust-worker-gender-biases.html</u>

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.