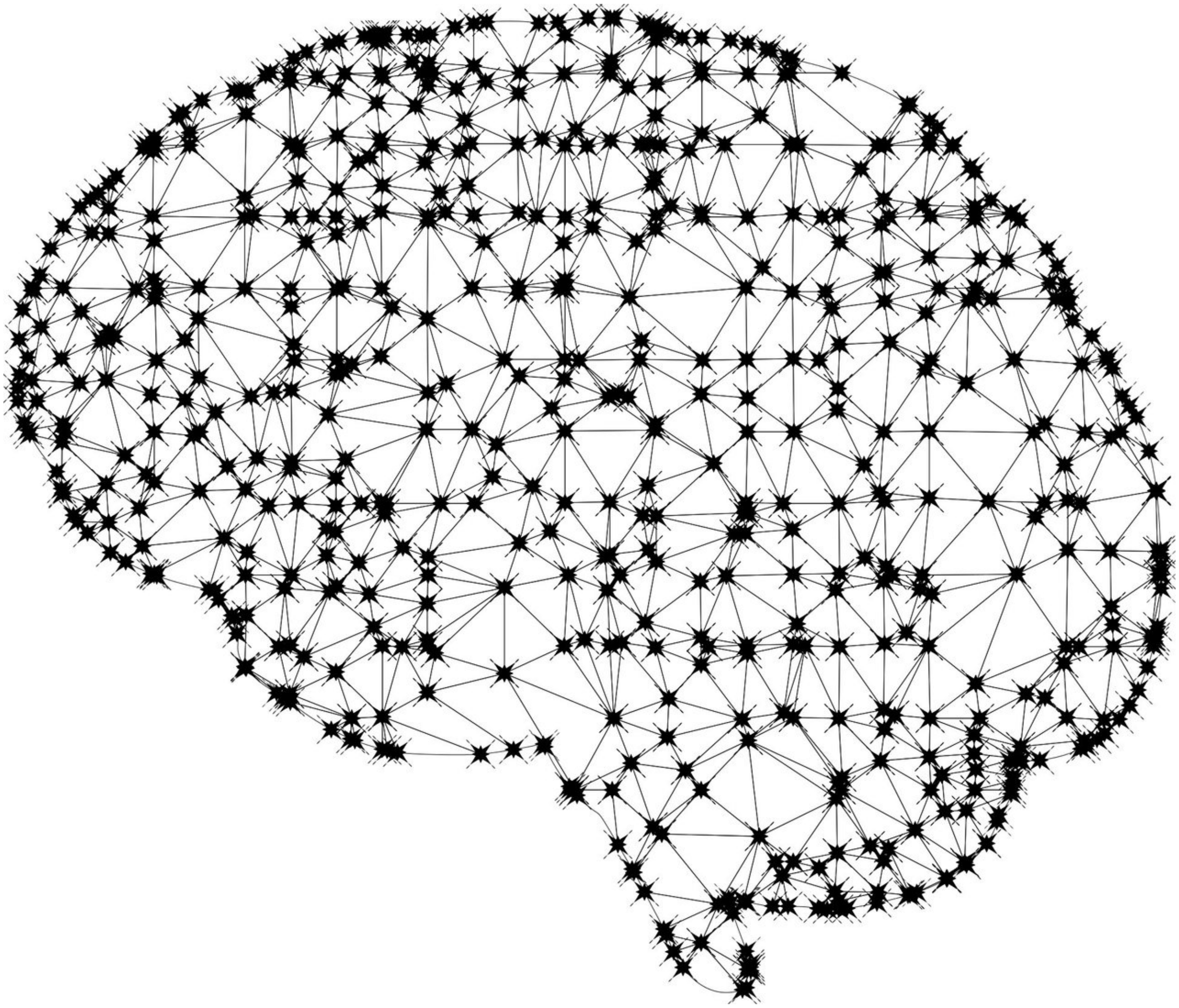
How to make AI less biased

November 16 2018



Credit: CC0 Public Domain

With machine learning systems now being used to determine everything from stock prices to medical diagnoses, it's never been more important to look at how they arrive at decisions.

A new approach out of MIT demonstrates that the main culprit is not just the algorithms themselves, but how the data itself is collected.

"Computer scientists are often quick to say that the way to make these systems less biased is to simply design better algorithms," says lead author Irene Chen, a Ph.D. student who wrote the paper with MIT professor David Sontag and postdoctoral associate Fredrik D. Johansson. "But algorithms are only as good as the data they're using, and our research shows that you can often make a bigger difference with better data."

Looking at specific examples, researchers were able to both identify potential causes for differences in accuracies and quantify each factor's individual impact on the data. They then showed how changing the way they collected data could reduce each type of bias while still maintaining the same level of predictive accuracy.

"We view this as a toolbox for helping machine learning engineers figure out what questions to ask of their data in order to diagnose why their systems may be making unfair predictions," says Sontag.

Chen says that one of the biggest misconceptions is that more data is always better. Getting more participants doesn't necessarily help, since drawing from the exact same population often leads to the same subgroups being under-represented. Even the popular image database ImageNet, with its many millions of images, has been shown to be biased towards the Northern Hemisphere.

According to Sontag, often the key thing is to go out and get more data from those under-represented groups. For example, the team looked at an income-prediction system and found that it was twice as likely to misclassify female employees as low-income and male employees as high-income. They found that if they had increased the dataset by a factor of 10, those mistakes would happen 40 percent less often.

In another dataset, the researchers found that a system's ability to predict intensive care unit (ICU) mortality was less accurate for Asian patients. Existing approaches for reducing discrimination would basically just make the non-Asian predictions less accurate, which is problematic when you're talking about settings like healthcare that can quite literally be life-or-death.

Chen says that their approach allows them to look at a dataset and determine how many more participants from different populations are needed to improve accuracy for the group with lower accuracy while still preserving accuracy for the group with higher accuracy.

"We can plot trajectory curves to see what would happen if we added 2,000 more people versus 20,000, and from that figure out what size the dataset should be if we want to have the best of all worlds," says Chen. "With a more nuanced approach like this, hospitals and other institutions would be better equipped to do cost-benefit analyses to see if it would be useful to get more data."

You can also try to get additional kinds of data from your existing participants. However, that won't improve things either if the extra data isn't actually relevant, like statistics on people's height for a study about IQ. The question then becomes how to identify when and for whom you should collect more information.

One method is to identify clusters of patients with high disparities in accuracy. For ICU patients, a clustering methods on text called topic modeling showed that cardiac and cancer patients both had large racial differences in accuracy. This finding could suggest that more diagnostic tests for cardiac or cancer patients could reduce the racial differences in accuracy.

The team will present the paper in December at the annual conference on Neural Information Processing Systems (NIPS) in Montreal.

More information: Why Is My Classifier Discriminatory? arXiv:1805.12002 [stat.ML] arxiv.org/abs/1805.12002

Provided by Massachusetts Institute of Technology

Citation: How to make AI less biased (2018, November 16) retrieved 20 April 2024 from https://techxplore.com/news/2018-11-ai-biased.html