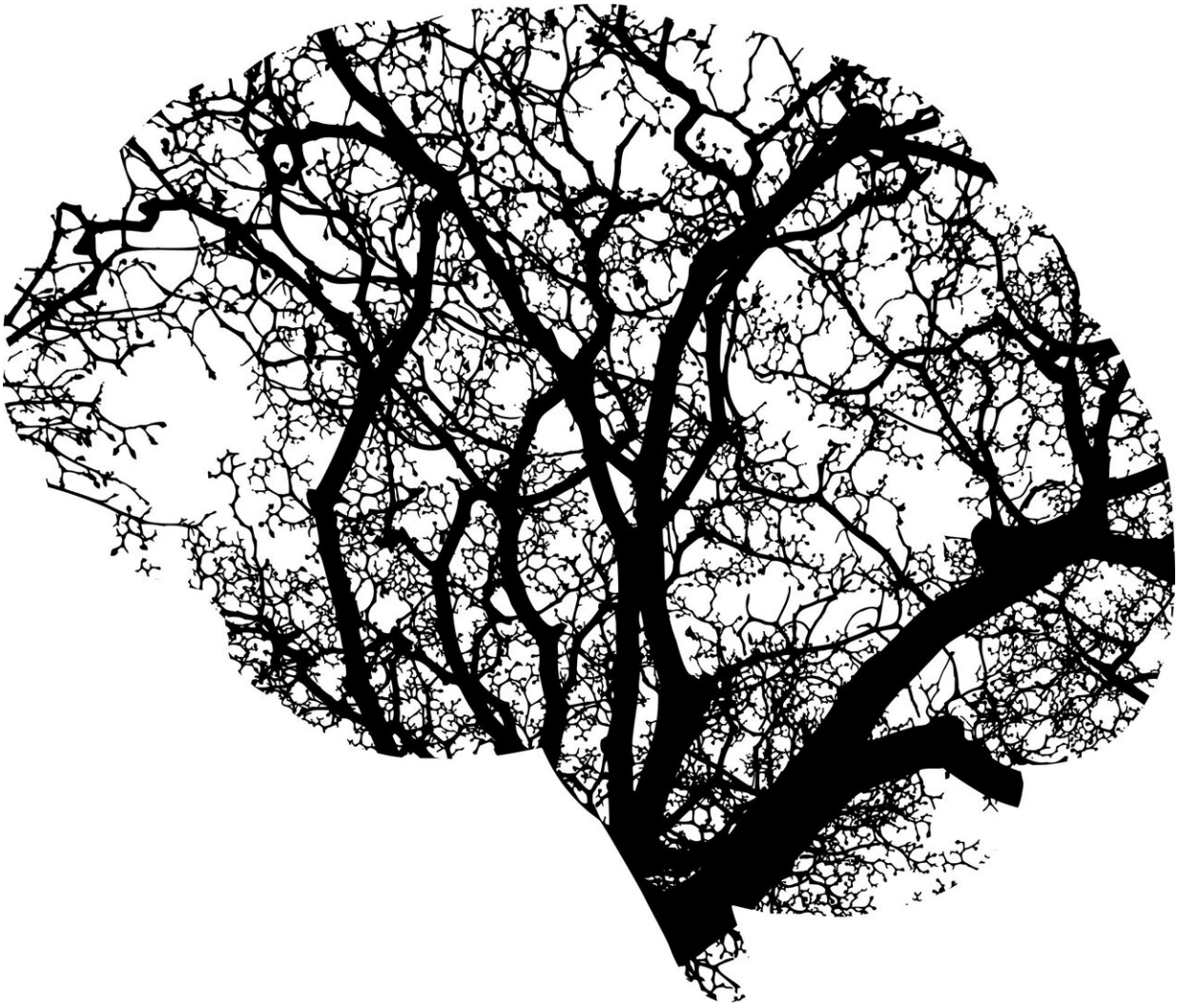
The blind use of AI in health care can lead to invisible discrimination

October 26 2023



Credit: CC0 Public Domain

Artificial intelligence can help health care systems under pressure allocate limited resources, but also lead to more unequal access. This is demonstrated by a research collaboration between the University of Copenhagen, Rigshospitalet and DTU that investigated whether AI can spot the risk of depression equally across different population segments. The research presents options for combing algorithms for bias prior to their deployment.

Artificial intelligence is making sure but steady headway in the health care system. MRI scans have already been made more effective by AI and Danish hospitals are now testing AI to make rapid emergency room diagnoses and better cancer diagnoses and therapies. But this is just the beginning.

On August 14, the Danish Minister of the Interior and Health, Sophie Løhde, stated that she envisions a future in which AI relieves beleaguered Danish health care system.

In hospitals and psychiatry, one of the tasks that AI is well suited to help improve is the allocation of limited funds through risk analyses and rankings which can ensure, for example, that therapies are provided to the patients where they can be most effective.

AI is already being used in other countries to assess who should receive treatment for depression. It is a development that could be on the way to a Danish mental health system under pressure.

Now, however, University of Copenhagen researchers are calling upon politicians for reflection, so that AI does not lead to more inequality or become an instrument for cold economic calculations. They point out that carelessness could make help a disservice.

"Artificial intelligence has great potential, but we need to be careful

because blindly implementing it can distort the health care system in new ways that are difficult to see, as the results can appear to be correct at first glance," says Melanie Ganz from the University of Copenhagen's Department of Computer Science and Rigshospitalet.

Invisible discrimination

In a new research article, she, along with her co-authors, documents how hidden biases sneak into an algorithm designed to calculate the risk of depression.

Together with colleagues from the Technical University of Denmark (DTU), the researchers developed the algorithm based upon algorithm design already being used in health care systems. Based on actual diagnoses of depression, the algorithm predicts people's risk of developing depression.

"In other countries, it is becoming more and more common to look at how to detect and prevent depression at an early stage. In the US, for example, AI is increasingly being used by private insurers to prioritize resources, a development that will likely come to Denmark in the near future. The question is, how fair will the foundation for such a prioritization actually be," says co-author Sune Holm from the Department of Food and Resource Economics.

The researchers used depression as a case study to investigate how we can evaluate the algorithms that we use both within the health care system and elsewhere in society, so that we can identify and adjust problems in time and make algorithms fairer before they are used.

"The right algorithms, if properly trained, can become tremendous assets for any municipality with limited resources. But our research shows that if machine learning isn't well managed, it can skew access to treatment,

so that some groups are overlooked or even left out," says Melanie Ganz.

The study shows that the algorithm has an easier time spotting the risk of depression in certain groups of citizens than in others based on the variables they are trained in—e.g. education, gender, ethnicity and a number of other variables. Indeed, the algorithm's ability to identify the risk of depression varied by up to 15% between different groups.

"This means that even a region or municipality, which in good faith introduces an algorithm to help allocate treatment options, can distort any such health care effort," says Melanie Ganz.

The algorithm can be a measurable success because it allocates resources to those who are in actual need. But at the same time, it could have hidden biases that exclude or deprioritize certain groups, without it being visible to those who are managing it.

At worst, AI systems can become an instrument of cold calculation. The choice of certain algorithms could be used to conceal the prioritization of resources for certain societal groups over others.

Tool to ensure fair algorithms

Sune Holm points out that AI also presents some fundamental ethical dilemmas.

"If we begin using these systems, it will be important to clarify who is responsible for prioritizing resources and individual therapeutic regimens should they be the result of algorithms. Furthermore, it could be difficult for a doctor to explain to a patient why a decision was made if the algorithm itself is incomprehensible," says Sune Holm.

While the research contributes theoretically to an area of machine

learning that deals with algorithmic discrimination across groups, the methods are also a robust tool for checking up on the quality of algorithm fairness.

"The methods we've developed can be used as a concrete prescription to evaluate the fairness of algorithms before they are used in, for example, municipalities and regions. In this way, we hope that the research can contribute to having the right tools in place for when the algorithms really make their entry into this area," says Melanie Ganz.

"Both politicians and citizens must be aware not only of the benefits, but also the pitfalls associated with the use of AI. So, one can be critical instead of just 'swallowing the pill' without further ado," says Sune Holm.

He believes there may be a need to ensure that the use of an algorithm has a documented positive effect on patients before investing in implementing it. For example, it should be clear how it can add value to the clinical practice that it is a part of.

The study is published in the journal *2023 ACM Conference on Fairness, Accountability, and Transparency*.

More information: Eike Petersen et al, On (assessing) the fairness of risk score models, *2023 ACM Conference on Fairness, Accountability, and Transparency* (2023). DOI: 10.1145/3593013.3594045

Provided by University of Copenhagen

discrimination.html