AI has potential to revolutionise health care—but we must first confront the risk of algorithmic bias

May 4 2023, by Mangor PedersenMangor Pedersen

Credit: Unsplash/CC0 Public Domain

Artificial Intelligence (AI) is moving fast and will become an important support tool in clinical care. Research suggests AI algorithms can accurately detect melanomas and predict future breast cancers.

But before AI can be integrated into routine clinical use, we must address the challenge of algorithmic bias. AI algorithms may have inherent biases that could lead to discrimination and privacy issues. AI systems could also be making decisions without the required oversight or human input.

An example of the potentially harmful effects of AI comes from an international project which aims to use AI to save lives by developing breakthrough medical treatments. In an experiment, the team reversed their "good" AI model to create options for a new AI model to do "harm".

In less than six hours of training, the reversed AI algorithm generated tens of thousands of potential chemical warfare agents, with many more dangerous than current warfare agents. This is an extreme example concerning chemical compounds, but it serves as a wake-up call to evaluate AI's known and conceivably unknowable ethical consequences.

AI in clinical care

In medicine, we deal with people's most private data and often life-changing decisions. Robust AI ethics frameworks are imperative.

The Australian Epilepsy Project aims to improve people's lives and make clinical care more widely available. Based on advanced brain imaging, genetic and cognitive information from thousands of people with epilepsy, we plan to use AI to answer currently unanswerable questions.

Will this person's seizures continue? Which medicine is most effective? Is brain surgery a viable treatment option? These are fundamental questions that modern medicine struggles to address.

As the AI lead of this project, my main concern is that AI is moving fast and regulatory oversight is minimal. These issues are why we recently established an ethical framework for using AI as a clinical support tool. This framework intends to ensure our AI technologies are open, safe and trustworthy, while fostering inclusivity and fairness in clinical care.

So how do we implement AI ethics in medicine to reduce bias and retain control over algorithms? The computer science principle "garbage in, garbage out" applies to AI. Suppose we collect biased data from small samples. Our AI algorithms will likely be biased and not replicable in another clinical setting.

Examples of biases are not hard to find in contemporary AI models. Popular large language models (ChatGPT for example) and latent diffusion models (DALL-E and Stable Diffusion) show how explicit biases regarding gender, ethnicity and socioeconomic status can occur.

Researchers found that simple user prompts generate images perpetuating ethnic, gendered and class stereotypes. For example, a prompt for a doctor generates mostly images of male doctors, which is inconsistent with reality as about half of all doctors in OECD countries are female.

Safe implementation of medical AI

The solution to preventing bias and discrimination is not trivial. Enabling health equality and fostering inclusivity in clinical studies are likely among the primary solutions to combating biases in medical AI.

Encouragingly, the US Food and Drug Administration recently proposed making diversity mandatory in clinical trials. This proposal represents a move towards less biased and community-based clinical studies.

Another obstacle to progress is limited research funding. AI algorithms typically require substantial amounts of data, which can be expensive. It is crucial to establish enhanced funding mechanisms that provide researchers with the necessary resources to gather clinically relevant data appropriate for AI applications.

We also argue we should always know the inner workings of AI algorithms and understand how they reach their conclusions and recommendations. This concept is often referred to as "explainability" in AI. It relates to the idea that humans and machines must work together for optimal results.

We prefer to view the implementation of prediction in models as "augmented" rather than "artificial" intelligence—algorithms should be part of the process and the medical professions must remain in control of the decision making.

In addition to encouraging the use of explainable algorithms, we support transparent and open science. Scientists should publish details of AI models and their methodology to enhance transparency and reproducibility.

What do we need in Aotearoa New Zealand to ensure the safe implementation of AI in medical care? AI ethics concerns are primarily led by experts within the field. However, targeted AI regulations, such as the EU-based Artificial Intelligence Act have been proposed, addressing these ethical considerations.

The European AI law is welcomed and will protect people working within "safe AI". The UK government recently released their proactive approach to AI regulation, serving as a blueprint for other government responses to AI safety.

In Aotearoa, we argue for adopting a proactive rather than reactive stance to AI safety. It will establish an ethical framework for using AI in clinical care and other fields, yielding interpretable, secure and unbiased AI. Consequently, our confidence will grow that this powerful technology benefits society while safeguarding it from harm.

Provided by The Conversation