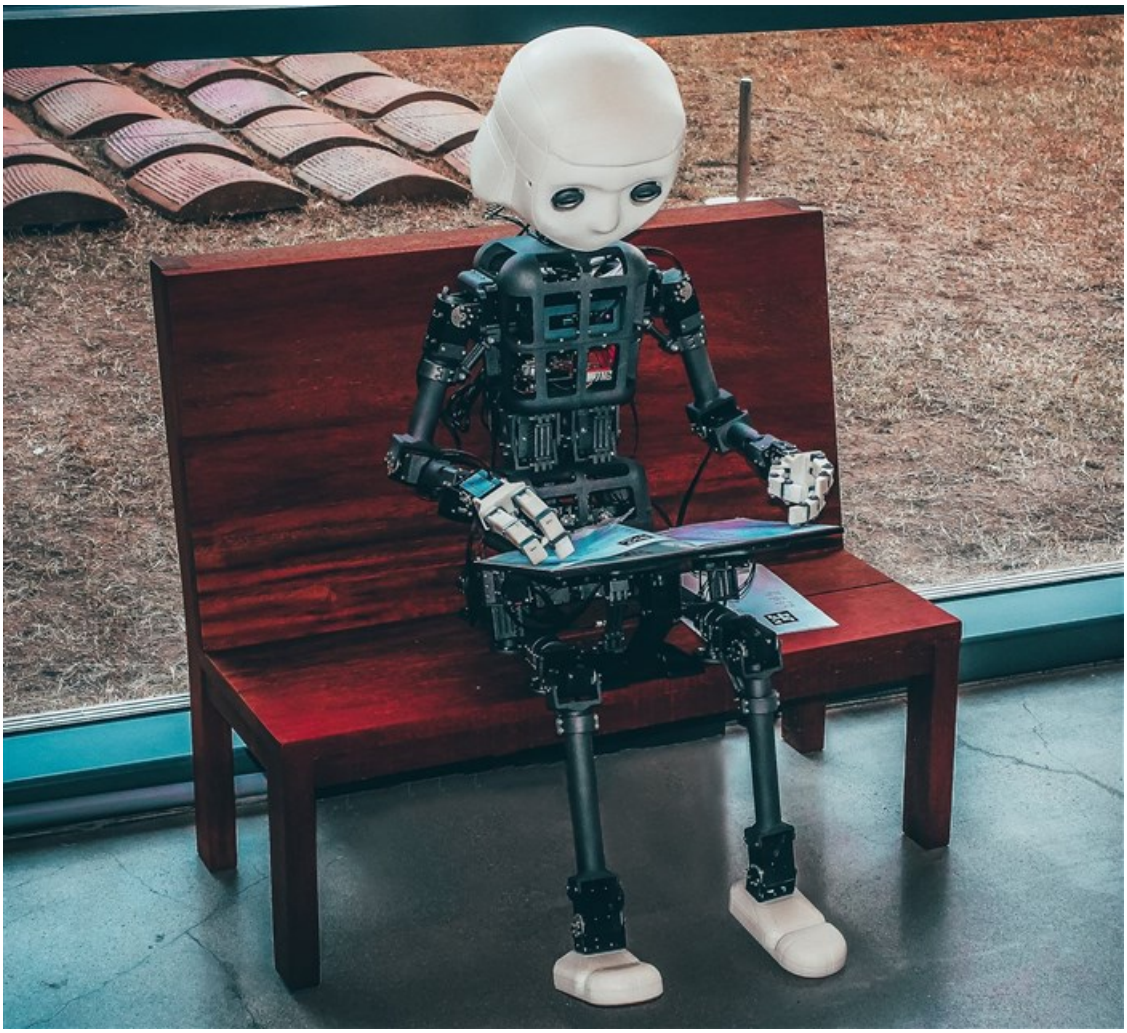
BIZCOMMUNITY

Defining what's ethical in artificial intelligence needs input from Africans

By Mary Carman, Benjamin Rosman

29 Nov 2021

Artificial intelligence (AI) was once the stuff of science fiction. But it's becoming widespread. It is used in mobile phone technology and motor vehicles. It powers tools for agriculture and healthcare.



Source: Unsplash

But concerns have emerged about the accountability of AI and related technologies like machine learning. In December 2020 a computer scientist, Timnit Gebru, was fired from Google's Ethical AI team. She had previously raised the alarm about the social effects of bias in AI technologies. For instance, in a 2018 paper Gebru and another researcher, Joy Buolamwini, had showed how facial recognition software was less accurate in identifying women and people of colour than white men. Biases in training data can have far-reaching and unintended effects.

There is already a substantial body of research about ethics in AI. This highlights the importance of principles to ensure technologies do not simply worsen biases or even introduce new social harms. As the UNESCO draft recommendation on the ethics of AI states:

> *"We need international and national policies and regulatory frameworks to ensure that these emerging technologies benefit humanity as a whole.*

”

In recent years, many frameworks and guidelines have been created that identify objectives and priorities for ethical AI.

This is certainly a step in the right direction. But it's also critical to look beyond technical solutions when addressing issues of bias or inclusivity. Biases can enter at the level of who frames the objectives and balances the priorities.

In a recent paper, we argue that inclusivity and diversity also need to be at the level of identifying values and defining frameworks of what counts as ethical AI in the first place. This is especially pertinent when considering the growth of AI research and machine learning across the African continent.

Context

Research and development of AI and machine learning technologies is growing in African countries. Programmes such as Data Science Africa, Data Science Nigeria, and the Deep Learning Indaba with its satellite IndabaX events, which have so far been held in 27 different African countries, illustrate the interest and human investment in the fields.

The potential of AI and related technologies to promote opportunities for growth, development and democratisation in Africa is a key driver of this research.

Yet very few African voices have so far been involved in the international ethical frameworks that aim to guide the research. This might not be a problem if the principles and values in those frameworks have universal application. But it's not clear that they do.

For instance, the European AI4People framework offers a synthesis of six other ethical frameworks. It identifies respect for autonomy as one of its key principles. This principle has been criticised within the applied ethical field of bioethics. It is seen as failing to do justice to the communitarian values common across Africa. These focus less on the individual and more on community, even requiring that exceptions are made to upholding such a principle to allow for effective interventions.

Challenges like these – or even acknowledgement that there could be such challenges – are largely absent from the discussions and frameworks for ethical AI.

Just like training data can entrench existing inequalities and injustices, so can failing to recognise the possibility of diverse sets of values that can vary across social, cultural and political contexts.

Unusable results

In addition, failing to take into account social, cultural and political contexts can mean that even a seemingly perfect ethical technical solution can be ineffective or misguided once implemented.

For machine learning to be effective at making useful predictions, any learning system needs access to training data. This involves samples of the data of interest: inputs in the form of multiple features or measurements, and outputs which are the labels scientists want to predict. In most cases, both these features and labels require human knowledge of the problem. But a failure to correctly account for the local context could result in underperforming systems.

For example, mobile phone call records have been used to estimate population sizes before and after disasters. However, vulnerable populations are less likely to have access to mobile devices. So, this kind of approach could yield results that aren't useful.

Similarly, computer vision technologies for identifying different kinds of structures in an area will likely underperform where different construction materials are used. In both of these cases, as we and other colleagues discuss in another recent paper, not accounting for regional differences may have profound effects on anything from the delivery of disaster aid, to the performance of autonomous systems.

Going forward

AI technologies must not simply worsen or incorporate the problematic aspects of current human societies.

Being sensitive to and inclusive of different contexts is vital for designing effective technical solutions. It is equally important not to assume that values are universal. Those developing AI need to start including people of different backgrounds: not just in the technical aspects of designing data sets and the like but also in defining the values that can be called upon to frame and set objectives and priorities.

ABOUT THE AUTHOR

Mary Carman is a lecturer in philosophy at, University of the Witwatersrand. Benjamin Rosman is an associate professor in the School of Computer Science and Applied Mathematics, University of the Witwatersrand.