In addition to the list of companies and governments that tried to address the issue of fairness in the use of AI, Google has listed some tips on checking for unfair biases on their website, as well as links to their own research papers on this topic.
PHOTO: REUTERS

For people to trust AI, build fairness into it

Algorithms learn from data given to them: if the data is biased, the results will be biased, so all those involved in the system need to be responsible. **BY DAVID DE CREMER**

NOT long ago, professional networking platform LinkedIn published a fairness toolkit, which is an open source software library that other companies can use to measure fairness in their own AI models.

This is an addition to the list of companies and governments that have tried to address the issue of fairness in the use of AI. Google lists some tips on checking for unfair biases on their website, as well as links to their own research papers on this topic.

Singapore has a Model AI Governance Framework that advises on how to translate fairness and transparency into practice, for example, making AI policies known to stakeholders. The European Union has the Ethics Guidelines for Trustworthy AI.

Some people may still look at the increasing use of AI with uneasiness. Think of the saga in which algorithms were used to predict the results of A-levels students in the United Kingdom based on how well they scored historically in their secondary schools. Many students' results were downgraded, particularly those from poorer schools. The use of AI, meant to reduce teachers' bias in predicting the students' results, created a new bias.

Algorithms learn from the data given to them. Hence, if the data is biased, the results produced by the algorithm will be biased too. In an incident this year, Detroit police wrongfully arrested an African American man as an AI facial recognition mistakenly identified him as a robbery suspect.

With these incidents, will end-users trust any algorithm's outputs to be fair? Companies using AI thus need to build trust. Research has shown that when people perceive that they are being treated fairly, they are more likely to cooperate and perform better. Similarly, people may be more acceptable of the algorithm's recommendations if they perceive the use of AI as fair.

How should governments, in their push for AI, and businesses who are using AI, convey to users that they are fair?

It is not simply a technological problem that requires factoring in certain features like accountability and interpretability in finding the right automation process. No, fairness is a social concept that humans use to decide how to interact with others, such that all parties will turn out for the better. The focus here is on a greater cause.

Hence if we want to see how fair an algorithm is, we need to go beyond its technical aspects and also take into account the social forces. And, one important force is to see how well AI is perceived as helping people collaborate for a greater cause.

INVOLVE EVERYONE

As a result, it should not just be the data engineers' and data scientists' teams that get to decide the fairness of the AI system. Governments, companies and organisations need to take into account as well the perceptions of stakeholders and end users. Do they feel like they have been treated fairly? What are their expectations of fairness?

As AI does not possess a moral character, it will also take a human (or humans) to evaluate whether the solution AI recommends is a fair one in the context of society. This reality requires that we need people who are aware of societal norms combined with strong ethical values.

Sure, humans have their own biases, but research has shown that they are less likely to be biased when it is about evaluating the decisions of others. In that respect, the development of an ethical compass cannot be emphasised enough both in schools and in workplaces.

REMEMBER TO BE HUMANE

As AI fairness is more complex than simply looking for a technical solution, the development of algorithms also requires a more humane approach. Rather than looking for an optimal and rationally employed algorithm, a compromise between the search for utility versus humanity will have to be sought for the use of AI to be accepted and considered fair.

Humans do not have a fixed productivity rate. We have short productivity bursts, we fare better in some environments than others, or we may take time as we mourn over family crises or adjust to new changes in our personal lives.

With the pandemic, many families have to adjust to living apart from their loved ones for long stretches of times. Some parents struggle with juggling working from home and their children's home-based learning. Some healthcare workers struggle with whether to put their children in a relative's house, or live with them and face the potential risk of spreading Covid-19 to them.

While AI can provide the most optimal allocation of resources and workflow management, there needs to be some buffer because humans will need help from time to time. As a society, we need to keep this in mind.

Ultimately, the verdict of whether an AI system is fair lies with the end user. It does not matter if the system has superior technical qualities, people will primarily judge its fairness based on how they perceive the algorithm is used to generate solutions and how those outcomes correspond with their values.

This means that data scientists and engineers must have these values in mind as they build the system. For example, if a company values the inclusivity of minority groups, then the data of these groups should be included and not treated as data anomalies.

For this reason people ultimately judge the company, not the algorithm, on whether it is fair. Indeed, in the eyes of the people, an algorithm is not a morally sensitive entity – it is a machine – and as such cannot be held responsible for the mathematical calculations it uses to generate solutions.

Instead, organisations and governments are led by humans and can be seen as responsible for how they make decisions, including their use of AI-systems. We need our leaders to be clear on their values and take into account the information needs and expectations of the people facing the results of the models' outputs. Only then will there be greater trust in AI.

▮ The writer is the Provost's Chair and professor in management and organisation at the National University of Singapore (NUS) Business School. He is also founder and director of the Centre on AI Technology for Humankind (AiTH). The opinions here are the writer's and do not represent the views and opinions of NUS.