# Artificial intelligence carries a huge upside. But potential harms need to be managed

By Alison Gillwald, Rachel Adams

7 Dec 2021

Artificial intelligence and machine learning have the potential to contribute to the resolution of some of the most intractable problems of our time. Examples include climate change and pandemics. But they have the capacity to cause harm too. And they can, if not used properly, perpetuate historical injustices and structural inequalities.



Source: Pexels

To mitigate against their potential harms, the world needs frameworks for the governance of data that are economically enabling and that preserve rights.

Artificial intelligence and machine learning operate on the basis of massive datasets from which algorithms are programmed to discern patterns. These patterns can be used to infer new insights and also predict behaviour and outcomes. Increasingly, artificial intelligence and machine learning are being used to substitute human decisions with automated decision making on behalf of humans. This is often in areas which can have a significant impact on peoples' lives. Take access to loans or even access into a country.

Yet it all happens in a black box that even the designer the algorithm may not have access to, so deciding what goes into the box is important.

The biggest datasets and algorithmic activity are generated by the global social networks that surveil our every action online. These datasets can be used to anticipate and mould our needs and desires.

Big technology firms, multilateral agencies and development banks have made much of the potential of artificial intelligence to advance economic growth and national development. And they're increasingly being used in social and economic applications as well as public decision-making, planning and resource allocation. These include guiding court judgments, selecting job applicants and assigning scholars to schooling systems.

The Covid-19 pandemic has also highlighted the enormous value of public data and the potential value of combining public and private data to deal with public health and disaster crises.

Yet, there is growing concern about the uneven distribution of both the opportunities and harms associated with artificial intelligence.

## The threats

The increasing use of artificial intelligence and machine learning in public decision making is raising critical issues around fairness and human rights.

In particular, how digital data are produced is being red-flagged. Datasets have some huge gaps. Certain people are rendered visible, underrepresented, and discriminated against as a result, in the way data are collected. The fact that most of world's population isn't connected to the internet and the global social networks that drive the new, data-driven economy means they simply don't exist.

Globally, artificial intelligence also poses a risk to the progress made toward gender equality. Stories abound of artificial intelligence systems being biased against women and gender minorities.

What's more, artificial intelligence systems may rely on assumptions and data that exclude or misrepresent groups that already face multiple and intersecting forms of discrimination. This often results in outcomes that reflect and reinforce gendered, racial, and ableist inequalities and biases.

These systems are not adequately subject to the kind of rigorous accountability and regulation needed to mitigate the risks they pose to society.

So significant is this threat that several international forums have emerged that are committed to the development of "good", "ethical" and "responsible" artificial intelligence.

But most of the initiatives present technical solutions to social and political problems. This means they are being developed outside a human rights frame. They are also largely initiatives of the global North, with limited multistakeholder participation from the global South.

## A right-based approach

There are rights-based data frameworks which inform artificial intelligence development. These include the European Union's General Data Protection Regulation. But they tend to focus primarily on first-generation or fundamental rights, such as privacy. Privacy is broadly conceived of as an individualised right. It may not always be the chief value in more communitarian-centred societies.

The Covid-19 pandemic has highlighted the need for data to be regulated in the collective interest or common good. This does not mean that the right to privacy
needs to be foregone.

Collective interest also pertains to the governance of data in the context of identifiable groups or communities where the potential consequence of individual identification results in the exposure of collective identity.

The literature and practice of data governance has predominantly been viewed and undertaken from this negative regulatory perspective. In other words, with a focus on compliance with data protection and cybersecurity and penalties for breaches.

This is a necessary condition for just artificial intelligence. But it's not sufficient. There are many areas of data governance that require positive intervention. Examples include enabling access to data, its usability and integrity if it is to deal with issues of inclusion, equality, redress, and social justice.

These are issues that can be understood as second and third generation, social and economic rights.

## AI that respects human rights

To address these issues, a new global project is being launched on the side-lines of the Summit for Democracy.

The summit represents an international forum to advance commitments in support of democracy and human rights. Its objective is to assess the progress being made by countries in advancing artificial intelligence that respects human rights and democratic values.

The project is known as the Global Index on Responsible AI. It is being led by the African digital think tank, Research ICT Africa, and an independent Data 4 Development network.

Governments and the international community have started to respond to the global call for responsible artificial intelligence. In 2019, 42 countries signed up to the OECD principles on Trustworthy AI. This commits them to ensuring that AI systems are safe, fair, and trustworthy.

Most recently, Unesco developed Recommendation on Ethics in AI was adopted by its 41st General Assembly. The recommendation centres on the protection of fundamental rights and freedoms, environmental sustainability, and diversity.

The Global Index addresses the need for an inclusive, measurable standard that complements the rapidly evolving understanding of what responsible artificial intelligence means in practice. It also encourages and tracks the implementation of governance principles by relevant actors.

The Global Index will track the implementation of responsible AI principles in over 120 countries. An international network of independent researchers will be established to assess the extent to which the principles are being applied. It will also collect primary and secondary data on key indicators of responsible artificial intelligence.

This will equip governments, civil society, researchers, and other stakeholders with the key evidence they need to uphold responsible use principles in the development and implementation of artificial intelligence systems. The evidence will also be used to:

- meet development and human rights obligations,
- build capacity for responsible AI around the world, and
- deepen international co-operation.

The public and other interested stakeholders will be given an opportunity to help shape the design and reach of the Index

which will be developed consciously through a global South lens.

Its development represents an important opportunity for experts from the African continent, and the Global South, to be at the forefront of shaping the new global agenda on the responsible use and development of artificial intelligence.

## ABOUT THE AUTHOR

Alison Gillwald: adjunct professor, Nelson Mandela School of Public Governance, University of Cape Town. Rachel Adams: doctoral supervisor, University of Cape Town.

For more, visit: https://www.bizcommunity.com