Artificial Intelligence, Bias, and the Sustainable Development Goals

Olivia Abbey, University of Virginia, United States of America, oca2px@virginia.edu

Abstract

Artificial intelligence is a rapidly developing global field of technology. If utilised properly, AI can increase productivity, eliminate human bias and error, and aid in achieving the UN's global sustainable development goals, particularly in facilitating better and more accessible healthcare, education, and economic growth. . However, if the data AI is trained on is reflective of systematic bias, it can perpetuate discrimination and deepen inequality. If AI's bias is not addressed before it spreads to all facets of global society, it could have disastrous results. This is a global problem that requires an innovative, global solution. The UN needs to create awareness about the potential dangers of AI bias and create global standard datasets to assist with the training of artificial intelligence algorithms.

Introduction

Artificial intelligence takes in vast amounts of data and detects patterns, adjusting itself with every new piece of information it takes in. If utilised properly, AI can reduce human error, increase productivity, and solve complex problems. However, if the data AI is fed reflects societal bias, AI can, and has, discriminated against people groups based on gender, race, location, and other characteristics.

This is a rapidly developing field. Many companies are starting to invest in research and adopt some AI into their models. The global artificial intelligence market is expected to double annually for the next ten years (GrandView Research, 2023). More than one third of all organisations have adopted AI in some capacity, ranging from driving, hiring, medicine, and national security organisations (IBM, 2022).

What is Artificial Intelligence?

Artificial intelligence is a rapidly developing field with huge potential. Although there is no clear definition of artificial intelligence it refers to machines that are capable of human level judgement, reasoning and execution of tasks (West, 2022). AI is able to absorb huge quantities of information much faster than any human can do and process and combing data or real world stimuli to identify complex patterns and respond accordingly (Hunt, 2014). Artificial intelligence is unique because of its ability to do deep learning, meaning it uses brain artificial neural networks to tweak its outcomes with each new piece of information, or "learn."

SDG and AI's Potential for Good

If used correctly, artificial intelligence has the potential to eliminate human error, make certain services faster and more accessible, and advance global progress towards the Sustainable Development Goals. Artificial intelligence might be particularly useful in the SGDs of health (3), education (4), and economic growth (8).

Artificial intelligence in healthcare is making great strides. There are many ways that AI is able to make the lives of individuals with health problems easier. There is AI to help blind people navigate new spaces, to help people with memory loss to enhance their capabilities, and even to predict when those with epilepsy are likely to seize (Rong, 2022). It has also proven to be capable of diagnosing diseases earlier and more accurately than experts through mechanisms such as gene analysis, signal processing, and medical imaging (Summerton, 2019). It has the potential to drastically reduce the cost of healthcare and bring expertise to low income countries and communities with few health care workers (Ciecierski-Holmes, 2022). AI can assist in repetitive tasks, freeing up time for doctors to help more patients.

AI can also make education more useful and accessible across the world. There have been many initiatives to harness the powers of AI to help students, teachers, and adults learn. For example, Uruguay has committed to integrating AI technology into their education policies to improve outcomes. This includes services such as a mathematics program that customises to the pace and learning style of students and then delivers performance metrics to teachers and helps them prepare assignments and exams accordingly (Pedro, 2019). Eneza Education is an AI platform in Kenya that provides materials, quizzes, and live Q&A with an AI chat using low data bandwidth short messaging service (EnezaEducation, 2023). With the help of artificial intelligence, more children and adults can receive personalised education experiences, even if they have limited or no access to a teacher.

The potential for economic growth in the upcoming years is huge. It is predicted that AI could

double annual global economic growth rates by 2035 by increasing labour productivity and innovation (Szczepanski, 2019). Economic growth will not necessarily just be reserved for developed countries, as AI has the potential to benefit many industries in low income countries, such as finances and agriculture. AI can help close the financial invisibility gap by creating more access to credit, assisting in retirement and savings plans, and doing risk assessments to provide loans to more people and small businesses. It is also able to help small farmers maximise their crop output and minimise the resources needed through weather and pest predictions and disease diagnostics (Savchuck, 2019). It can also help initiatives to eliminate poverty. For example, low income housing is being built for a quarter of the cost ten times faster through 3D printing and artificial intelligence modelling (Dharmaraj, 2018).

What is Artificial Intelligence Bias?

However, AI also has the potential to produce a lot of harm, whether it be through privacy violations, deepening the digital divide, social surveillance and automating discriminations. Algorithm bias is one of the most dangerous aspects of artificial intelligence.

Artificial intelligence bias occurs when the recommendation or predictions are discriminatory or misrepresentative of a group of people. The successes of AI rely on its ability to provide accurate and useful information, which is dependent on it being given objective and reliable data. Without careful selection of the datasets being fed into algorithms, AI has the potential to exacerbate many existing problems that people groups of different gender, race, disabilities, location, and other characteristics face. Data used to AI that is reflective of societal bias can make discrimination across the globe automated, undetectable, and greatly magnified.

As mentioned, artificial intelligence has so much potential because of its ability to learn which it is only able to do because of the large amounts of data available from the internet. Devices connected to the internet, ranging from heart monitors to computers, generate data at unprecedented volume, variety, and velocity, appropriately labelled "big data" (O'Leary, 2013). AI then sorts this data to make it useful to researchers, companies, and individuals and applies the data trends to its own frameworks. Because artificial intelligence uses huge amounts of data, it could, in theory, come very close to the objective truth. But even big data can be riddled with bias.

The digital divide, which the UN is working hard to address, means that certain communities and even countries are underrepresented in data sets. Additionally, even if data sets are accurately representative of reality, that reality may be a result of historic and current exclusion of communities. Even if certain inputs are excluded from datasets, such as race and gender, proxy characteristics as a result of systems might still result in discriminatory outputs (Savchuck, 2019).

The bias can come not just from the data itself but also from developers. The way they intrust AI to prioritise certain factors and ignore others may contribute to discriminatory results. These developers are typically white, rich, educated men, and AI may reflect their social standing (Zhang, 2021).

AI Bias and Sustainable Development Goals

Unless there are intentional steps to mitigate artificial intelligence bias, it may hinder the success of the UN's sustainable development goals globally. Already, there have been clear illustrations of AI bias worsening the inaccessibility of healthcare, education, and economic growth.

AI bias in healthcare can have deathly effects. Many medical studies have been conducted with sample populations that are not representative of the true populations. Approximately 80% of genetics data comes from Caucasians. Artificial intelligence risk scores of cardiovascular diseases and other genetic conditions are far more accurate for white people than other racial groups. This risk score heavily influences who is referred to for further medical examinations and support (Igoe, 2021). Historically and in the present, women tend to be underdiagnosed as symptoms are primarily reflective of men's experiences and their pain dismissed. This is reflected in data globally and would impact the ability of AI to give proper medical assistance to female patients.

There are many pitfalls for AI in education to reinforce and deepen academic disparities. For example, AI predicting retention rates for higher education overestimated white students graduation rates and underestimated latinos graduation rates. AI for evaluating communication competency rated German students much lower than human evaluation did and Chinese students much higher (Baker, 2021). E-learning platforms that have the potential to increase education accessibility, may also reinforce stereotypes in their recommended courses (Boratto, 2019). The A level work given to artificial intelligence to develop

grading algorithms reflects teachers' bias and has been proven to penalise disadvantaged students (Ofqual, 2019).

A huge area of concern for artificial intelligence and economic growth is hiring bias. While AI can help automate processes such as resume review, allowing for more productivity in workplaces, it can also reinforce discrimination. Algorithms that seek to ignore race and gender may in fact reinforce gender and racial bias (Drage, 2022). Marginalised groups may be further marginalised by AI, increasing the amount of people who are unable to be financial contributors to their economy.

Policy recommendations / conclusions

Although there is no one solution for eliminating artificial intelligence bias, there are several things the UN can do to be a part of the solution.

First, the EU should work to create best development practices for AI. They should do this in partnership with working groups from the Institute of Electrical and Electronic Engineers, National Institute of Science and Technology, European Group on Ethics in Science and New Technology, and other intercontinental organisations. These groups should research AI bias and provide guidance to countries on how to regulate the use of artificial intelligence, eliminate bias, and utilise this technology for advancing the sustainable development goals and eliminating discrimination. Together, they should establish best development practices in AI and facilitate accountability and transparency in countries developing AI.

Additionally, the UN should work to compile datasets from their member states and individual companies, as well as through their own research and expertise. These standard datasets need to be high quality, representative, explainable, and transparent data that is obtained in a way that has the consent of the data subjects. It is a difficult and time consuming task of compiling best practice data sets and requires many rounds of verification that it is accurate and does not reinforce biases, but, if done correctly, it can be used to train AI in ways that are beneficial towards the SDGs and not harmful. Training AI with clean data allows AI to better identify and address "dirty data", data that is biassed or leading to inaccurate results. These data sets can be distributed to people and member states looking to use AI to advance the SGD's and eliminate bias.

The UN needs to be at the forefront of tackling artificial intelligence bias. Artificial intelligence has huge potential for creating progress but it also has many possible pitfalls. Eliminating artificial intelligence bias will require innovation, investment, and global collaboration. The UN has the power and resources to work with other science and technology groups to create best practice standards and to distribute data sets that mitigate bias in this emerging technology.

References

Baker, R. S., & Hawn, A. (2021). Algorithmic bias in education. International Journal of Artificial Intelligence inEducation, 1-41.

Boratto, L., Fenu, G., & Marras, M. (2019). The effect of algorithmic bias on recommender systems for massive open online courses. In *Advances in Information Retrieval: 41st European Conference onIR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part I 41* (pp. 457-472). Springer International Publishing.

Ciecierski-Holmes, T., Singh, R., Axt, M., Brenner, S., & Barteit, S. (2022). Artificial intelligence for strengthening healthcare systems in low-and middle-income countries: a systematic scoping review. *npj DigitalMedicine*, *5*(1), 162.

Dharmaraj, V., & Vijayanand, C. (2018). Artificial intelligence(AI) in agriculture. *International Journal of Current Microbiology and Applied Sciences*, *7*(12), 2122-2128.

Drage, E., & Mackereth, K. (2022). Does AI Debias Recruitment? Race, Gender, and AI's "Eradication ofDifference". *Philosophy & technology*, *35*(4), 89. https://doi.org/10.1007/s13347-022-00543-1

Eneza Education. (2023). *Spreading education everywhere.*Eneza Education. Retrieved March 5, 2023, from https://enezaeducation.com/

Grand View Research. (2023). Artificial Intelligence Market Size & Share Analysis Report 2030. Retrieved March5, 2023, from https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market

Hunt, E. B. (2014). Artificial intelligence. Academic Press.IBM. (2022, May). *IBM Global AI Adoption Index 2022*. Retrieved March 5, 2023, from https://www.ibm.com/downloads/cas/GVAGA3JP

Ofaqual. (2019). *GCSE, AS & A level Summer Report 2019*. Gov.uk. Retrieved March 5, 2023, from https://assets.publishing.service.gov.uk/governmen t/uploads/system/uploads/attachment_data/file/1 052706/Levelling_Up_WP_HRES.pdf

O'Leary, D. E. (2013). Artificial intelligence and big data. *IEEE intelligent systems*, *28*(2), 96-99.

Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). Artificial intelligence in education: Challenges and opportunities for sustainable development.

Rong, G., Mendez, A., Assi, E. B., Zhao, B., & Sawan, M. (2020). Artificial intelligence in healthcare: review andprediction case studies. *Engineering*, *6*(3), 291-301.

Savchuk, K., Spiess, J., & Gillis, T. (2019, October 28). *Big Dataand racial bias: Can that ghost be removed from themachine?* Stanford Graduate School of Business.Retrieved March 5, 2023, from https://www.gsb.stanford.edu/insights/big-data-ra cial-bias-can-ghost-be-removed-machine

Summerton, N., & Cansdale, M. (2019). Artificial intelligence and diagnosis in general practice. *British Journal ofGeneral Practice*, *69*(684), 324-325.

Szczepanski, M. (2019). Economic impacts of artificialintelligence (AI).

West, D. (2022, March 9). *What is Artificial Intelligence?* Brookings. Retrieved March 5, 2023, from https://www.brookings.edu/research/what-is-artifi cial-intelligence/

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., ... & Perrault, R. (2021). The AI index 2021 annual report. *arXiv preprint arXiv:2103.06312*.