ZIMUN XI

The Cost of Innovation:

Navigating the Ethical Responsibilities of Technological Advancements and Societal Change for Inclusive, Sustainable Development in a Globalised World

Human Rights Council

Resolving the Discriminatory Implications of Algorithmic Bias and Ensuring Ethical AI development and Equitable Outcomes

Committee: Human Rights Council

Issue: Resolving the discriminatory implications of Algorithmic Bias and ensuring ethical AI development and equitable outcomes

Student Officer: Ruvarashe Jakanani

Position: Chair

INTRODUCTION

All Machine learning Artificial Intelligence is taught how to solve problems by responding to certain stimuli in the environment until it finds a suitable method to solve a problem through trial and error. Algorithmic bias occurs when a set of outcomes to solve a problem is favored by an algorithm. These can come as a result of a large consumer demand which shifts the algorithm in an outcomes favor and can be in the form of socioeconomic, gender, religious or racial bias. Overcoming AI algorithmic bias is considered by many to require the same skill needed to overcome social biases such as gender stereotypes and racial profiling, and would require data experts with a mastery in the craft to handle it.

DEFINITION OF KEY TERMS

- **Algorithmic Bias**: "algorithmic bias describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others" – (Florida State University Libraries, Noble)
- **Algorithm**: "a procedure for solving a mathematical problem in a finite number of steps that frequently involves repetition of an operation." – (Merriam Webster dictionary)
- **Bias**: "an inclination of temperament or outlook" – (Merriam Webster Dictionary)
- **Discrimination**: "prejudiced or prejudicial outlook, action, or treatment" – (Merriam Webster Dictionary)
- **Artificial Intelligence**: "the capability of computer systems or algorithms to imitate intelligent human behaviour" – (Merriam Webster Dictionary)

- **Machine learning:** "a computational method that is a subfield of artificial intelligence and that enables a computer to learn to perform tasks by analysing a large dataset without being explicitly programmed." – (Merriam Webster Dictionary)

BACKGROUND ON THE ISSUE

Algorithmic bias first appeared in several medical practitioners where racial bias was experienced between patients seeking treatments in 2019. The concept was introduced in the 1980s after reports at St George's university noted that an algorithm created by Dr Geoffrey Franglen would deduct points on student applications based on race and gender, where students with non Caucasian names where deducted up to q5 points and women up to 3 points in average, impacting the student diversity greatly.

More research was held by prominent figures and learners such as Safiya Noble, Joy Buolamwini, Ruha Benjamin and Cathy O'Neil. In 2019 the issue was reintroduced by Tristan Panch, Heather Mattie and Rifat Atun under healthcare after concerns about social class, gender and race were brought up with regards to quality of healthcare and less and more effective treatments. The introduction of AI into medicine in particular has raised concerns about black boxed systems, where the design and construction of healthcare systems and common algorithms is kept secret, implying that it would be easy to implement algorithmic bias into such systems. As noted by Pasquale – "we can observe its inputs and outputs, but we cannot tell how one becomes the other" (Pasquale, 2015 p3)

The issue grew momentum in 2016 where investigative journalism from ProPublica showed that various forms of software used in the United States of America showed a bias which would often predict future criminals as African Americans. Moreover scholar Joy Buolamwini researched into the aspect of Amazon facial recognition being more prone to error when used for dark skinned women. The issue mainly impacts non – Caucasian groups and causes problems for them when gaining essential services such as banking, medical care and even basic job opportunities when legal connotations of systemic racism are introduced. It is also important to note that machine learning algorithms are fed information by the users and take note of priority data, meaning that such algorithms have learnt harmful racial and

social stereotypes from human behavioural characteristics which now impacts what should be fair, adequate and streamlined practices in the modern age.

The international community has concerns over this AI overdevelopment and resulting in hyper biased systems world wide which sum fear may not be reversed. It is important to study this issue and understand the basic functions of AI and its interactions with humans for it to be a tool used to better our world.

CURRENT CONTEXT

Recently, an article from The Conversation discusses the United States bringing AI into social services. Concerns about whether AI is a good tool to bring into child protection were raised after several cases of AI having inappropriate answers to prompts were brought up. In 2019, a case was filed against HireVue, the AI hiring tool used by several companies to hire employees. It tended to favour certain facial features as well as gender orientation. 40% of all African AI companies were created in the last ten years, and many face challenges with program sophistication due to a lack of adequate infrastructure and algorithmic bias further hinders the development of such programs.

 A study by Buolamwiniand Gabri held in 2018 showed that out of 1270 participants from Finland, Sweden, Rwanda, South Africa and Iceland still showed that dark skinned women were more likely to be miscategorised than fair skinned males. Particularly in this story, the error rate for light skinned men was 0.8%, whereas the error rate for dark skinned women was 34.7%.

MAJOR COUNTRIES AND ORGANIZATION INVOLVED

World economic forum: key players advocating for good AI practices which are sustainable, promote rules and regulations hoping to lessen AI bias

IEEE: researches and develops upon challenges in AI along with its researchers

The OECD: The main UN group involved in this issue, mainly informs the public about the ethical implications of AI and its use,has noted that it has several privacy concerns and that more management is needed in the handling of these tools.

The United States Accountability Office: an office was formed here specialising particularly in AI and its uses.

TIMELINE OF KEY EVENTS

1970s - Dr Geoffrey Fraglen of St George's Hospital Medical school begins developing an algorithm to be used for student applications into the university.
It was found that the initial way of accepting applicants was crucial to the university therefore Fraglen developed an algorithm to assist in the speed and accuracy of the process

1979: Fraglen's algorithm is completed.

1982: All applicants are being screened by the algorithm
The algorithm was deemed reliable as it agreed with the approval panel 90% to 95% of the time.

1980s - The algorithm created by Dr Geeoffrey Fraglen is investigated
Concern is raised by university students and teachers who have noted a lack of diversity among applicants who have passed.

1980s - First incident of algorithmic bias in AI found and documented
Studies held by the university found that dark skinned women were more likely to have points deducted from their initial application than light skinned men. Furthermore, applicants with non european names were more likely to have points deducted than applicants with European names.

December 1986 - The algorithm is taken to the UK Commision for racial equality

Two senior lecturers at St George's medical university suspected that the algorithm was used as a form of racial prejudice. St George's was found guilty in the case.

2014 - Reuters interviews Amazon employees about the Amazon hiring process

The system would reportedly rate each of the applications a number of stars out of five, notable favouring the top results.

2015 - The way the bias in the algorithm was created is revealed

It was found that the algorithm used by amazon had a gender bias due to the input information being applications from high ranking male workers in the industry.

2015 - It is revealed that the Amazon algorithm tended to discriminate against any application with the word 'women's'

For example, 'rugby team' would be favoured over 'women's rugby team' in the algorithm.

2016 - Other cases in algorithmic bias pop up

This includes the ProPublica situation mentioned earlier in this report`1

2017 - present - International concern raised over AI bias

RELEVANT UN RESOLUTIONS, TREATIES, & EVENTS

Australian Human Rights Commission - Using Artificial Intelligence to make decisions: addressing the problem of algorithmic bias

- Made in 2020

European Union Agency for Fundamental Rights - Bias in algorithms-artificial intelligence and discrimination

- Made in Vienna in 2022

ITU - [United Nations Activities on Artificial Intelligence](#)

- Made in 2021

UNIDIR - [Algorithmic Bias and the weaponisation of the increasingly autonomous technologies](#)

UN News - [Bias, racism and lies: facing up to the unwanted consequences on AI](#)

- 30 December 2020

UNU - [The Dual Faces Of Algorithmic Bias: avoidable and unavoidable discrimination](#)

- 31 January 2024

PREVIOUS ATTEMPT TO SOLVE THE ISSUE

One attempt was the IBM AI Fairness 360 toolkit which was an open source library used to detect biases in AI programs. It is used to support developers in mitigating algorithmic bias when spotted as well as the necessary tools to develop new programs without such biases. Furthermore , Microsoft's fairlean is also a toolkit designed to assess the fairness of AI models as well as mitigate and understand biases in their program. To add on to this, MIT Media Labs justice league is a leading source of research on AI accountability and is a constant advocate for AI awareness. The partnership on AI is like the MIT Media Labs, and involves NGOs, companies and academic institutions all committed to researching and raising awareness on AI issues and accountability.

POSSIBLE SOLUTIONS

Some possible solutions that can be backed by research are as follows:

1. Discussing that the input data used to initially develop these algorithms is the boundary data to the extreme. For instance, in the case of the Amazon AI application fiasco, the input data used from the algorithms came from real life leaders in the industry who mostly happened to be light skinned males. Instead of using real life data which may be subject to heavy bias, model input data could be developed where the UN issues a standard policy for algorithmic application input data that has equal portions for several ethnicities. This could be done easiest by including extreme data types for ethnicity and gender and allowing these to be the set boundaries of the algorithm.

    - E.g. extreme data - dark skinned lesbian woman to like skinned heterosexual man : these would become the boundaries of the algorithm and all other test data could flow in between which would help reduce error in categorising applicants in systems

2. Having humans watch over the use of these algorithms rather than to let them run on their own.

    - In this way, companies, businesses and other organisations can fix any error in the algorithm on hand and also assist in the application process. It is important to note that the individuals watching over these algorithms would need to be trained not to have a bias towards certain groups when watching over this algorithm and the application process

3. Discourage overdependence on AI

    - There is still a general lack of knowledge towards AI therefore it may be important to advocate for its use in moderation in certain fields.

WORKS CITED

Addressing AI Bias: Real-World Challenges and How to Solve Them
https://www.digitalocean.com/resources/articles/ai-bias

Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies - (2018)
https://unidir.org/files/publication/pdfs/algorithmic-bias-and-the-weaponization-of-increasingLy-autonomous-technologies-en-720.pdf

Algorithmic bias in artificial intelligence is a problem—And the root issue is power - (september to october 2023 issue)
https://www.sciencedirect.com/science/article/pii/S0029655423001288#:~:text=A

Articles on Algorithmic bias - (10 February 2025)
https://theconversation.com/ai-is-being-used-in-social-services-but-we-must-make-sure-it-doesnt-traumatise-clients-248555

BIAS IN ALGORITHMS ARTIFICIAL INTELLIGENCE AND DISCRIMINATION - (2022) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

Bias, racism and lies: facing up to the unwanted consequences of AI - (30 December 2020)
https://news.un.org/en/story/2020/12/1080192

Navigating algorithm bias in AI: ensuring fairness and trust in Africa- (24 October 2024)
https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2024.1486600/full

Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges, and Perspectives - (December 2022)
https://ieeexplore.ieee.org/document/9645355

Shedding light on AI bias with real world examples - (16 October 2023)
https://www.ibm.com/think/topics/shedding-light-on-ai-bias-with-real-world-examples

The Dual Faces of Algorithmic Bias — Avoidable and Unavoidable Discrimination - (31 January 2024)
https://unu.edu/article/dual-faces-algorithmic-bias-avoidable-and-unavoidable-discrimination

United Nations Activities on Artificial Intelligence (AI) - (2021)

https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2021-PDF-E.pdf

Untold History of AI: Algorithmic Bias Was Born in the 1980s

https://spectrum.ieee.org/untold-history-of-ai-the-birth-of-machine-bias

Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias -
(2020)

https://humanrights.gov.au/sites/default/files/document/publication/final_version_technical_p
Aper_addressing_the_problem_of_algorithmic_bias.pdf

What is algorithmic bias - (17 July 2023)

https://www.datacamp.com/blog/what-is-algorithmic-bias