International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

The Manifestation and Implications of Bias in Artificial Intelligence on Global Society

Aarav Singh

Student, Jayshree Periwal International School

Abstract
Although artificial intelligence (AI) is becoming more and more prevalent, biases from data, algorithms, and feedback loops might still exist. Marginalized groups may be disproportionately affected by these biases, which may result in biased outcomes in areas such as facial recognition, credit scoring, and hiring. Reduced trust, sustained societal inequality, and stifled innovation are some of the effects. It is critical to support transparent development processes, diverse development teams, and ethical oversight in AI design in order to lessen these effects. If these issues are resolved, AI will become a more equitable instrument for advancing society.

1. Introduction
Artificial Intelligence, coined as AI, has been an integral part of the present-day society. Imitating human intelligence through machines has revolutionised the social infrastructure and become a highly dependable sector. However, the impending growth of reliability in artificial intelligence has raised concerns about the bias and its embryonic repercussions on discrete democratic groups.
This research paper revolves around whether AI could be prejudiced and if so, what negative implications it could enthral over mankind. However, we would scrutinise how biases could manifest in AI to get a generalised notion, further delving into the repercussions in various societal domains and on individuals.

2. Manifestation of Bias in AI
Artificial Intelligence could be prone to biases which mostly encompasses the fact that it is man-made. Biases in AI could emerge from diverse sources, each giving rise to potentially skewed information and decisions.

2.1 Biased Training Data
The most commonly referred to prejudice is biased training data. This indicates that during the AI system learning process, the knowledge that is imputed could be one-sided. If the training specifics are not particularly diverse and embedded from limited sources, then the AI model may develop prevailing subjectivity based on the learned wisdom. An example of this could be language bias. Natural Language Processing models embedded with historical shreds of evidence could contain gender or racial stereotypes and perpetuate these biases.

2.2 Feedback Loop Bias
Another commonly occurring distortion is in AI systems which intercommunicate with surroundings or interactors and prosper a skewed specific established on the information they acquire from the feedback. Feedback looping is when the received data from interactions is itself biased. E-commerce platforms highly rely on recommendation systems based on user preference and interaction in the feedback loops.

One common form of bias in recommendation systems is popularity bias. The system tends to exhibit popular items favoured by users to attract newcomers. This dynamic creates a feedback loop where niche or less popular items may overshadow the precedence dispensed for other products. The bias establishes a need for more diversity displayed to the customers.

2.3 Algorithmic Bias

Occurring when algorithms are formulated in certain directions to display favouritism towards certain outcomes, algorithmic bias leads to embedded prejudice in mathematical structure, dissociated with the data they process. Commonly occurring, credit scoring algorithms might disadvantage certain zip codes or income brackets and may result in the denial of loans to creditworthy individuals from specific socioeconomic backgrounds. Credit scoring algorithm is a model that delves into the assessment of an individual's creditworthiness through in-depth financial analysis to predict the likelihood of credit repayment. A fault in such is a serious negative implication to the society. Similarly, facial recognition systems may be remotely efficacious for particular skin tones or even genders. A seminar study by researchers Joy Buolamwini and Timnit Gebru titled “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification” led to a discerning conclusion. Facial analysis algorithms display significant inaccuracies when identifying gender in images of darker skin. The study disclosed an error rate of 0.0% to 0.3% for lighter-skinned males compared to the highest of 20.8% to 34.7% for darker-skinned females. The algorithm displayed 0.8% to 7.4% error in the case of darker-skinned males and 0.0% to 0.12% for lighter-skinned females.

To recapitulate, the propensity of Artificial Intelligence to inherent biases in its design by humans highlights the necessity to overlook and deduce comprehensive strategies to address the tendentiousness.

3. Implications of AI Bias in the Global Society

Due to the expeditious growth of AI, the fact arises that it could be quite lethal in terms of perpetuating and amplifying biases. Nowadays, AI has become a highly relied-upon resource which increases the influence of biases.

3.1 Erosion of Public Trust

The current growth in technology and the availability of the current population has increased the reliance on AI. This implies that biased AI can erode public knowledge and communication, leading to pessimism and reluctance to welcome AI solutions. This could further undermine the growth in impactful areas such as healthcare, or research and development. Furthermore, customers relying on IT companies may become sceptical of their findings due to biases. The undermining of trust could hinder the development of efficient resource allocation and decision-making processes, diminishing their potential exploits.

3.2 Exacerbation of Social Inequalities

Considering the biases in the world itself, based on gender and cast, a skewed AI may just heighten the discriminations already occurring. This is solely based on where AI systems produce results to cause an unfair advantage to certain groups based on their race or gender. In an employment system, an AI discriminatory towards certain demographic groups, could undermine their qualifications and provide preference for other genders. This may reduce job opportunities and strike against the gender equality norm. For example, if an AI system is biased against women, then their qualifications could be undervalued in preference to other groups and may reduce their job opportunities.

3.3 Hindrance to Innovation and Adoption of New Ideas

Additionally, biases could hinder the development and adoption of new ideologies. It may prove to be more sceptical to accept fresh ideas and explore new possibilities. This can lead to a tapering outlook and a trepidation to venture into uncharted territories. Biased AI could generate barriers to entry for innovations, particularly those whose ideas do not align with the biased AI structure. This may cause difficulty in accepting new ideas from certain marginalized parties, even though their proposal is superior. Furthermore, it may be unfair to them as the recognition the groups deserve cannot be attained.

4. Influence on Different Age Groups and Individuals

Biases in AI not only influence society as a whole but also impact certain age groups and individuals in different ways.

4.1 Impact on Education (5-18 years)

AI technologies are highly incorporated into the education of children between 5 to 18 years. They immensely rely on AI as a basic means to solve their problems and create presentations and research work. A bias in such a case could disrupt their knowledge and intelligence, proving as an obstacle to future growth. They may perpetuate those discriminations and skewed information, potentially impacting their educational outcomes. Moreover, a bias of AI based on personalized learning could prove to be difficult for certain individuals who do not fit the profile where the AI favours certain learning modalities.

4.2 Impact on the Workforce (18-35 years)

Progressing to the major working population of 18 to 35 years, AI could cause major issues. AI in this field majorly focuses on hiring and employment by screening and filtering suitable candidates based on qualification and performance evaluation. A biased system towards certain genders, ages, or educational backgrounds may limit the actual potential and acceptance of some people, further hindering their career progression. As this is the age their pivotal moment and holds the key to shaping their future, biases could forestall their potential and success in the workforce.

4.3 Impact on the Elderly Population

The elderly population may also be adversely affected by AI bias. Neglect of adequate age-related factors by AI systems in healthcare may provide inappropriate recommendations or overlook critical health issues. This can lead to reduced quality of care and marginalization of older individuals in society.

5. Strategies To Mitigate AI Bias

Reducing biases in AI requires unprejudiced and multi-faceted approaches that address stages of system development.

5.1 Diversifying Training Data

Commencing with an approach to diverse the feedback and information provided in forming the AI. It should not be inclined towards a particular distinction and serve as a representation of society as a whole. Ensuring diversity in datasets of AI can diminish the inherent bias it comprises.

5.2 Implementing Ethical Guidelines and Regulations

To boot, forming laws and regulations like supervising data collection and governance could prove to diminish the biases. It could even be required for AI developers to exhibit the liability and transparency of their algorithm along with ensuring that AI systems comply with anti-discriminatory laws, further proving to be a safeguard against unfair practices. An oversee of development, conducted by regulatory bodies, could ensure unprejudiced outcomes, proving to be key to mitigating the occurring biases.

5.3 Inclusive AI Development Teams

All-inclusive teams, be it any gender, ethnicity, age, culture or comprising different experiences, would allow to bring numerous prospective designs and amendments to the current systems. Diversity would serve as a playmaker to cover blind spots that homogenous teams might overlook or perceive with a biased approach. By involving a range of voices in AI development, the systems would produce more accurate and equitable judgements.

5.4 Continous Monitoring and Evaluation

Be it anything, but continuous monitoring and evaluation of AI systems is prominent to detect biases that may occur at any stage. Regular check can identify uncommon patterns in AI outputs. Establishing feedback loops allows organizations to update algorithms based on findings. This ongoing oversight proves as a holistic approach to depleting the biases of AI and maintaining their integrity over time.

6. Conclusion

The proclivity of Artificial Intelligence to inherit biases in its design by humans highlights the necessity to develop comprehensive strategies to address these prejudices. Biased AI can erode public trust, exacerbate social inequalities, hinder innovation, and negatively impact individuals across different age groups. Addressing AI bias, therefore, requires a holistic approach that includes diversifying training data, implementing ethical guidelines, promoting inclusive development practices, continuous monitoring, and increasing public awareness.

7. Bibliography

1. Engler, Alex, et al. “Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms.” *Brookings*, 27 June 2023, www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/. Accessed 03 Oct. 2024.
2. Buolamwini, Joy, and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *PMLR,* 21 Jan. 2018, proceedings.mlr.press/v81/buolamwini18a.html. Accessed 03 Oct. 2024.
3. Padilha, Victor A., and Andre C.P.L.F. de Carvalho. “A Study of Biclustering Coherence Measures for Gene Expression Data | IEEE Conference Publication | IEEE Xplore.” *IEEE*, ieeexplore.ieee.org/document/8575671/. Accessed 03 Oct. 2024.
4. “Home.” *AI Now Institute*, 1 Aug. 2024, ainowinstitute.org/. Accessed 03 Oct. 2024.
5. “Google Responsible AI Practices.” *Google AI*, ai.google/responsibility/responsible-ai-practices/. Accessed 03 Oct. 2024.
6. Buolamwini, Joy, and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *PMLR,* 21 Jan. 2018, proceedings.mlr.press/v81/buolamwini18a.html. Accessed 03 Oct. 2024.