Biases in Artificial Intelligence Applications Affecting Human Life: A Review

Ravindra Kumar

Abstract: *The introduction of Artificial Intelligence has improved operations in almost every sector, industry, and part of human life. The use of AI has been vital in the department of justice, recruitment by organizations, facial recognition by police, and school admissions. The aim of introducing AI algorithms in various fields was to reduce human bias in decision-making. Despite the progress, there are ethical concerns that the AI algorithms also exhibit biases. The main reason behind the claim is because human developers are in charge of training data used by the algorithms. There are areas where the issue of biases affects human life directly and can do damages to a person, physically or emotionally. Some examples are college admissions, recruitment, administration of justice at the courts, public benefits systems, police, public safety, and healthcare. There are high chances that the development process introduced biases in artificial intelligence algorithms, knowingly or unknowingly, during any area mentioned above. The paper provides background knowledge on AI bias and possible solutions to solve the problem.*

Keywords: Biases In AI, AI, Artificial Intelligence, social AI

I. INTRODUCTION

In today's technology world, most business entities and individuals have recognized the importance of Artificial intelligence. The recognition has led them to adopt AI in their operations. Despite the overwhelming benefits of AI, there are some concerns. There is concern on AI whether it will be a threat to humanity. Another concern is regarding job security, where most people feel that it will replace them. There is also concern that AI possesses human biases in decision-making associated with training data [1]. AI bias portrays loopholes in the output of the machine algorithm. The biases originate from the data used during the testing and development phases. Biasness in AI has been experienced in hiring, passing judgment in courts, making arrests, and facial recognition systems. In most instances, vision systems have identified minorities as criminals because of the mindset of the developer. Cognitive biases emanate from feelings towards a given individual because of group membership. Cognitive biases can be associated with the human history of decision-making. Psychologists have identified that there are 180 biases in humans. The biases impact our decision-making and problem-solving. Humans have developed AI algorithms, and we might introduce the element of bias unknowingly or during training [2]. The history of racism in America, for example, has contributed to the dominant racial discrimination. The established policies and procedures and their adoption in AI technology contribute toward practicing inequalities in criminal justice.

Lack of complete data is another type of bias which affects decision making in AI algorithm. When we obtain information, research findings, survey, or intelligence for data gathering from a specific population group, and they are not representative of the entire population; it creates a bias in the resulting algorithm. Relying upon and feeding such information to AI systems leads to biases in decision making. It is vital to remember that algorithm bias in AI can be evident in gender bias, racial prejudice, and discrimination by age. AI systems developed with in-built learning capabilities develop their biases from the interaction and information sharing between different parties.

II. LITERATURE SURVEY

There have been several instances where AI bias has been identified. Earlier in 2016, Microsoft developed an autoreply bot for Twitter. The AI system could learn from people's interactions, and it started responding to people with racist comments. The bot was considered racists and used comments which affected some individuals. Hence twitter avoided using the bot. Another example is the Amazon recruiting tool which was implemented in 2014. The introduction of the AI system was seen as the best approach to help recruiters in screening resumes for potential candidates to fill the vacancies. However, it was realized that the AI system was biased towards women [4]. The problem emanated from the training data used for the AI system. The data used contained recruitment history for the past ten years when men dominated the technology field. Hence the machine perceived that men were preferred over women in recruitment. Amazon stopped using the tool because of the biased results. The healthcare risk algorithm has also portrayed biases in decision-making. The system was developed to help in determining the patients that required critical care. The healthcare sector realized there was a problem when the machine prioritized white patients over black patients. The designers had used historical data of income between the two races during training [3]. Issues of racial discrimination placed whites at high income over blacks; hence they spent more on healthcare. The machine perceived that because whites spent more on healthcare, they were worth being given priority. Facial recognition technologies have also been paused based on portraying biases.

54

According to the Association for Computing Machinery (ACM), government and private institutions should not use facial recognition systems. According to them, the systems portray biases on ethnic, racial, gender, and other human features [3].

Because of those biases, people in a given demographic group are affected by the decisions made by the facial recognition systems. The biases in the system have portrayed a negative impact on society. Countries like the United States and China have deployed surveillance cameras that monitor citizens without their consent. Such surveillance has been opposed because of the homogeneous composition of engineers involved in developing the systems. Marginalized communities have repotted various instances where they have been wrongly profiled because of their race leading to distrust of the government.

III. PROPOSED METHODS

The rapid implementation of AI in businesses and government institutions shows the dependence of technology in society. It is vital to ensure that the systems which are designed to help in decision-making offer minimal biases. It is the responsibility of developers to ensure that the training data are well tested to eliminate the transfer of human bias to the algorithm. Despite the dependence on the AI system, human involvement should also be present in situations where human beliefs are central. Government can play an essential role in eliminating or controlling the bias by setting accountability. Just like there are laws in the share market for insider trading, HIPPA in healthcare, and **some other examples**, the government can by law require companies to preserve train and test data for any artificial intelligence application that affects human lives in any way. Of course, there is a need for discussion at larger scales about ethics, rules, and boundaries. Another way of eliminating bias can be to open test data for public scrutiny or audit. With AI applications in use, the test data to train the model can be posted on public folders so that if someone suspects bias in the application, they can review test data and raise the concerns, if any. Bounty programs to find bias, just like bounty programs to find bugs in the application, can also be a great way to determine the bias introduced in AI application, knowingly or unknowingly.

IV. CONCLUSION

The rapid implementation of AI in businesses and government institutions shows the dependence of technology in society. It is vital to ensure that the systems which are designed to help in decision-making offer minimal biases. It is the responsibility of developers to ensure that the training data are well tested to eliminate the transfer of human bias to the algorithm. Despite the dependence on the AI system, human involvement should also be present in situations where human beliefs are central.

REFERENCES

1. Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings Institute: Washington, DC, USA.
2. Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. The handbook of evolutionary psychology, 1-20.
3. Kantarci, A. (2020). Bias in AI: What it is, Types & Examples, How & Tools to fix it. Retrieved from https://research.aimultiple.com/ai-bias/
4. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 469-481).

AUTHORS PROFILE

Ravindra, is a software engineer, technologist, and researcher in the healthcare IT, population health management, clinical software applications product management, and Implementation space. He has over a decade of proven experience in leading and making global teams spanning various functions like Technical, Solutions Architecture, Technical Alliances, Development, and Quality Engineer. He is passionate about Artificial Intelligence, Machine Learning, and Computer Vision. He wants to use Artificial Intelligence, Machine Learning, and Computer Vision for social goods and healthcare, and youth welfare worldwide; hence, all his research is focused in these areas. Ravindra participates actively in designing and implementing new AI algorithms and open source projects that impact how humans can be healthier, How we can mitigate health and safety risks, and how youth can stay away from sadness. Ravindra is a keen reader and participates in deep learning and AI meetups and forums. Ravindra is also active in agile coaching and ensures teams adopt agile practices in the right way. He also loves to train people in quality engineering practices, including RPA(Robotic Process Automation), Test Automation, and Agile engineering, programming for AI and ML.