Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

38

Bias Mitigation in AI Technology

Kruti Pratik Kotak

Research Scholar, Computer Science, Surendrangar University, Wadhwan

Abstract

Artificial Intelligence (AI) is revolutionizing the way decisions and predictions are made across various domains, from business and finance to government services and healthcare. Its potential to enhance efficiency, productivity, and economic growth is evident, with PwC research estimating that AI could contribute $15.7 trillion to the global economy by 2030. However, the widespread adoption of AI also poses significant challenges, most notably the issue of bias. This research paper delves into the critical topic of bias mitigation in AI technology. It explores the implications of AI bias on decision-making, the limitations of technical solutions, and the broader strategies needed to address bias, making a case for the proactive involvement of business leaders and governance in shaping AI's future. This paper deals a comprehensive analysis of the challenges associated with bias in AI technology and highlight the critical role of business leaders and governance in shaping the future of AI. It will also underscore the importance of moving beyond technical solutions to address the broader dimensions of bias, fostering an equitable and inclusive AI-driven world.

Keywords: Artificial Intelligence, Bias Mitigation, Fairness, Ethics, Discrimination, Decision-making, AI Technology, Bias Challenges, Bias Mitigation Approaches, Bias Impact, Business Implications, Societal Implications, Case Studies, Best Practices, Future Directions.

1. Introduction

In recent years, Artificial Intelligence (AI) has gained substantial traction in businesses, particularly in processes like customer service, marketing, and sales. Its implementation is seen as a means to predict consumer behavior and gain a competitive edge. However, the increased reliance on AI has sparked concerns about human cognitive biases affecting AI-driven predictions. AI systems are not immune to errors, and biased data inputs can result in biased outputs. For instance, even when no explicit bias is programmed into AI systems, gender bias was observed in loan decisions and career stem advertisements. These biases raise concerns about automated decisions potentially leading to discrimination and undesirable outcomes. Interestingly, it is noted that technological flaws are more common than human errors in many firms.

The application of AI in sales to boost revenues has garnered global attention. Automation powered by machine learning, deep learning, information retrieval, and natural language processing offers innovative solutions, enhanced customization, profit optimization, and business transformation. However, AI biases have been observed in sales, where the algorithms generate business leads but may not effectively identify high lifetime value prospects. Transparency is recognized as a critical factor in AI deep learning systems, as it ensures privacy and justifiability of decisions. The "black box" problem can arise, hindering explainability and contributing to biases within the AI system.

The prevalence of AI biases extends beyond businesses, affecting areas such as court decisions and healthcare. Biased AI systems, like those causing gender bias at Apple or African American defendant bias in COMPAS, are expected to increase, potentially exploiting vulnerable populations. The awareness of AI biases has grown, particularly in fields like healthcare and business management, where incorrect projections can lead to reduced customer satisfaction and loyalty, impacting equity and profitability.

However, AI-driven decision-making introduces algorithmic biases, with limited research in this area. These biases can have profound effects on businesses, especially in decision-making contexts. Addressing automation biases concerning race, gender, credit scores, and face recognition is crucial, highlighting issues with virtual assistants, robotics, and algorithmic recommendations. As AI solutions are incorporated into firms, a critical approach is essential. The study raises questions about the nature of AI biases and how to address them, aiming to shed light on these biases and their mitigation within business systems. The research provides a novel perspective on understanding and addressing AI biases to mitigate potential risks.

2. Definitions of AI

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think, reason, and make decisions like a human. AI systems can perform tasks that typically require human intelligence, such as problem-solving, learning, perception, language understanding, and decision-making.

AI is the field of computer science dedicated to creating systems that can perform tasks that would typically require human intelligence. These tasks include understanding natural language, recognizing patterns, making decisions, and solving complex problems.

Table 1 Definitions of AI.

Description	Refs
The ability to reason, solve problems, learn, and integrate multiple human skills like perception, cognition, memory, language, or planning refers AI intelligence.	Kar et al. (2022)
AI systems use mathematical models to derive inferences from data, increases transparency and humans get answers to the questions like ’what’, ’how’ and ’why’ to bring the benefits to the business	Kar et al. (2022)
AI has evolved in the firms from being a just adopted technology to powering routine decision-making processes in all the domains	Kar et al. (2022); Morande (2022)
AI techniques able to increase the knowledge of employees in the firms by allowing them to comprehend and conquer complex situations more effectively and facilitates the decision-making process by offering several alternative choices	Malik et al. (2021)
The proficiency of a machine leveraged by AI to full fill the customer expectations and increases the operational efficiency in the	Kushwaha et al. (2021)

organization	
The term machine learning is an artificial or computational intelligence technique describes a machine's capacity to learn and carry out a process given an objective and specific training tasks to accomplish the goal	Votto et al. (2021); Garg et al. (2021)
AI is defined as systems which mimic cognitive abilities commonly associated with human characteristics such as learning, speech, and problem solving	Dwivedi et al. (2021)

3. Biases in AI systems

In cases where data inputs are inherently biased, the resulting outputs of AI systems are likely to be skewed. For instance, the controversial use of AI tools by Amazon to assess and rate job applicants resulted in significant gender discrimination. Furthermore, AI-related errors have been observed in insurance companies, where automated premium calculations factored in religion rather than gender. Such biases have manifested in dynamic pricing and targeted discounts, illustrating how algorithmic bias can seep into automated systems. Numerous studies highlight the adverse consequences of human biases that arise alongside technological advancements.

Cognitive biases, which affect decision-making processes, find their way into AI and robotic creations, thereby influencing a wide array of contexts. With the growing prevalence of human-like technology in consumer markets and marketing practices, it becomes imperative to understand how inadvertent human biases can permeate the design of artificial intelligence. The transfer of human biases to AI systems is often a result of the programming and coding process, which inadvertently introduces elements of racism and discrimination.

In essence, AI biases can be attributed to the transmission of human biases into machine learning algorithms, arising from preconceived assumptions during algorithm development or biased training data sets. These challenges underscore the non-trivial nature of algorithmic bias issues, emphasizing the need for education among marketers and consumers.

However, the introduction of bias in AI and robotic systems represents a complex and intricate challenge. This study delineates AI bias in the realms of psychology and behavioral economics, categorizing it into observable (identifiable through the analysis of extensive customer data, such as frequently purchased products and availability) and unobservable biases (associated with complex, difficult-to-detect big data, requiring advanced research skills). The multifaceted nature of AI biases spans across various industries, demanding a comprehensive approach to address both observable and unobservable risks (Table 2).

Table 2 AI bias in various industries.

Algorithm bias	Firm
Adobe software blocked certain tiers of customers when purchasing software	Adobe
Dynamic rideshare pricing has a bias that suggests customers have to pay higher price increases	Lyft
AI bias occurs on FB and allows advertisers to target marketing/job ads to specific gender, racial, and religious minority backgrounds.	Facebook(FB)
Fintech failures start with payment processes and revenue sharing partnerships	Banker
When a customer started chatting with the Tay chatbot about a racist comment, the voice assistant started retyping the sentence instead of responding to the customer's request.	Microsoft
The data-driven bias stems from Nikon's new product regarding Asian faces and HP Media Smart computers, which have skin color issues with facial recognition.	Nikon
This is one of the important new products (devices) in clinical management for monitoring oxygen levels during the pandemic. This created a bias (racial bias) in which people with darker skin tones were	Pulse oximeter

less accurate than people with lighter skin.

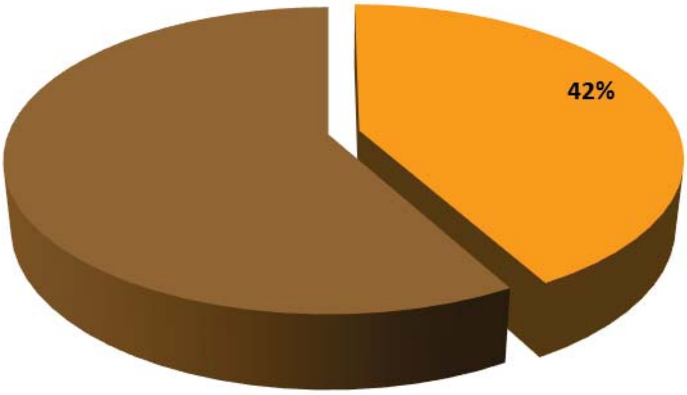
4. How does biased AI impact business?

Addressing bias in AI is both a moral imperative and a strategic necessity for businesses. Biased AI systems yield inaccurate and discriminatory outcomes for specific segments of the population. They can unfairly distribute opportunities, resources, and information, potentially infringing on civil liberties, compromising individual safety, providing unequal services, and negatively impacting well-being by being derogatory or offensive.

The ramifications for businesses are profound. Biased AI systems tarnish reputations, erode consumer trust, and undermine future market prospects. Tech companies acknowledge this risk, with Microsoft highlighting the threat of reputational damage or legal liability from biased AI systems in a report to the US Securities and Exchange Commission.

Biased AI systems often necessitate significant alterations or even abandonment, incurring substantial costs in terms of employee time and other resources. For instance, Amazon recalled an AI-driven hiring tool in 2018 that was biased against women, causing internal conflicts and increased calls for ethical practices.

On a broader societal scale, biased AI systems reinforce and magnify existing societal discrimination, leading to economic inefficiencies and market losses. Governments are responding with regulatory initiatives, implying potential penalties for companies that neglect bias mitigation.

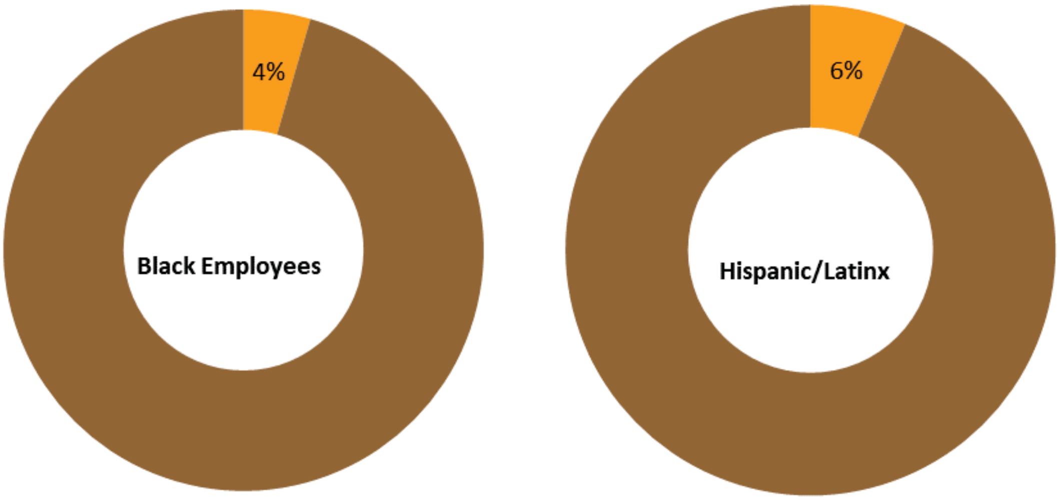


A 2019 DataRobot report revealed that 42% of organizations involved in AI systems production and use are "very" to "extremely" concerned about the reputational damage arising from media coverage of biased AI, highlighting the urgency of addressing this issue.

5. Why are AI systems biased?

AI systems are inherently biased, reflecting the society in which they are created. They are human constructs, essentially classification technologies deeply influenced by the context of their development. The people behind AI system development play a pivotal role, infusing their perspectives, knowledge, values, and priorities into these systems, thereby shaping their outcomes.

Regrettably, the tech industry's demographics compound this issue. Large-scale AI system developers, predominantly male and white, fail to represent the diverse makeup of society. Women account for only 26% of the computing workforce today, a lower percentage than in 1960, with nearly half leaving the tech field, a departure rate more than double that of men. Additionally, less than one fifth of AI researchers and professors are women. Racial diversity is also sorely lacking, with only 4.5% of employees at a leading US tech company being Black, and 6.3% being Hispanic/Latinx, statistics echoed across the tech sector.



Furthermore, AI bias unwittingly infiltrates the entire development and utilization of machine learning AI systems, particularly during data generation, collection, labeling, and management, as well as in algorithm design and evaluation. Despite noble intentions, this bias results in inaccurate predictions and discriminatory outputs, posing significant risks to individuals and businesses. Detailed insights into the pathways through which bias permeates datasets, algorithms, and AI system usage are essential to address this complex issue effectively.

6. Conclusion:

The rapid integration of Artificial Intelligence (AI) into various business processes, including sales, marketing, and customer service, offers exciting opportunities for companies to enhance decision-making, increase efficiency, and drive profitability. However, this proliferation of AI has brought to the forefront a significant and complex challenge: the presence of biases in AI systems. These biases, which can manifest in data inputs and algorithmic decisions, have far-reaching consequences, including discrimination and unintended outcomes.

The findings of this study shed light on the pervasive issue of AI bias, illustrating its presence not only in business but also across diverse industries, including finance, social media, and healthcare. Biased AI systems can amplify societal inequalities and have a profound impact on businesses, eroding trust, tarnishing reputations, and incurring substantial financial and legal liabilities.

Understanding the origins of AI bias is essential. Human biases, both implicit and explicit, find their way into AI systems during data collection, labeling, algorithm development, and decision-making processes. Additionally, the demographic makeup of the tech industry, dominated by white males, contributes to the perpetuation of biases in AI systems.

Addressing AI biases is both a moral imperative and a strategic necessity for businesses. By doing so, they can improve fairness, build trust with consumers, and position themselves for future success. It is vital for organizations to recognize that combating AI bias requires a multi-faceted approach, including diversity in the tech industry, careful data handling, and the development of transparent AI systems that can be explained and justified.

As businesses continue to rely on AI to drive innovation and gain a competitive edge, the need to confront and mitigate AI bias becomes increasingly urgent. Firms must engage in proactive measures, encourage diversity, and invest in ethical AI development to ensure that the transformative power of AI is harnessed responsibly and equitably. This study serves as a call to action, emphasizing the importance of understanding and addressing AI biases to mitigate risks and create a more inclusive and just business environment.

Reference

- Bach, A. K. P., Norgaard, T. M., Brok, J. C., & van Berkel, N. (2023, April). "If I Had All the Time in the World": Ophthalmologists' Perceptions of Anchoring Bias Mitigation in Clinical AI Support. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems: 1-14.
- Cranfield, J. A., Eales, J. S., Hertel, T. W., & Preckel, P. V. (2003). Model selection when estimating and predicting consumer demands using international, cross section data. Empirical economics, 28(2), 353–364
- Derman-Sparks, L. (2016). Guide for selecting anti-bias children's books. Teaching for Change Books.
- Gonzales, R. M. D., & Hargreaves, C. A. (2022). How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system. International Journal of Information Management Data Insights, 2(2), Article 100130.
- Harmon, P., & King, D. (1985). Expert systems: Artificial intelligence in business. John Wiley & Sons, Inc..
- Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49(1), 30–50.
- Leavy, S., O'Sullivan, B., & Siapera, E. (2020). Data, power and bias in artificial intelligence. arXiv preprint arXiv:2008.07341.
- Malik, N., Tripathi, S. N., Kar, A. K., & Gupta, S. (2021). Impact of artificial intelligence on employees working in industry 4.0 led organizations. International Journal of Manpower.
- Nilsson, N. J. (1982). Principles of artificial intelligence. Springer Science & Business Media.
- Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a standard for identifying and managing bias in artificial intelligence. NIST special publication, 1270(10.6028).

- Smith, Genevieve and Ishita Rustagi (2020). Mitigating Bias in Artificial Intelligence: An Equity Fluent Leadership Playbook. Berkeley: Haas School of Business, University of California
- Timmons, A. C., Duong, J. B., Simo Fiallo, N., Lee, T., Vo, H. P. Q., Ahle, M. W., ... & Chaspari, T. (2023). A call to action on assessing and mitigating bias in artificial intelligence applications for mental health. Perspectives on Psychological Science, 18(5), 1062-1096.
- Varsha P.S. (2023). How can we manage biases in artificial intelligence systems – A systematic literature review. International Journal of Information Management Data Insights 3 (2023) 100165
- Winston, P. H. (1992). Artificial intelligence. Addison-Wesley Longman Publishing Co., Inc.