ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/goutham-sabbani/186

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 1

Research Article

Ethics and Bias in AI: Challenges and Solutions

Goutham Sabbani*

Citation: Sabbani G. Ethics and Bias in AI: Challenges and Solutions. *J Artif Intell Mach Learn & Data Sci 2023*, 1(1), 747-749. DOI: doi.org/10.51219/JAIMLD/goutham-sabbani/186

Received: 03 March, 2023; **Accepted:** 28 March, 2023; **Published:** 30 March, 2023

***Corresponding author:** Goutham Sabbani, MSc FinTech, UK

ABSTRACT

Artificial intelligence has both advantages and disadvantages; A 2018 quantitative impact study showed that facial recognition systems e have an error rate of up to 34.7% for people with darker skin tones compared to 0.8% for lighter-skinned individuals, underscoring significant biases5.

AI has evolved from simple rule-based systems to complex machine learning models like neural networks, leading to advanced capabilities and increased resilience across various sectors. However, this evolution has also intensified ethical concerns and biases in these systems.

We will talk about the origins and implications of ethical issues with artificial intelligence. Also, talk about the causes of biases in AI systems and examine case studies to understand the real-world impact of AI biases. Furthermore, we will delve into potential solutions for migrating these barriers, proposing strategies for creating more ethical and unbiased AI technologies. By addressing these critical issues, we aim to foster a deeper understanding of AI ethics and promote the development of fairer AI systems.

Keywords: AI Bias, Ethical AI, Machine Learning, Algorithmic Fairness, AI Transparency

Integration of AI in daily life has brought several changes, and it has become a cornerstone of modern society, changing various sectors like healthcare and transportation. The adoption of this technology has led to advancements in efficiency, accuracy, and overall quality of life. However, the influence of this technology has a dual impact, presenting both remarkable benefits and significant challenges. For instance, in terms of advancements, dot cos can use personalized treatment to improve outcomes but also minimize adverse effects, thereby enhancing the overall efficiency and patient experience[5].

On the flip side, AI concerns regarding policy ethical implications and displacement of jobs are prominent. One criteria area is the bias inherent in some AI systems-a significant bias in 2018 revealing facial recognition technology. The automation capabilities of AI have led to the removal of substantial numbers of jobs. They became equipped to perform tasks that we humans previously did. The displacement of jobs raises concerns and totally changes the job landscape[2].

Here is a line chart showing the exponential removal of jobs by AI

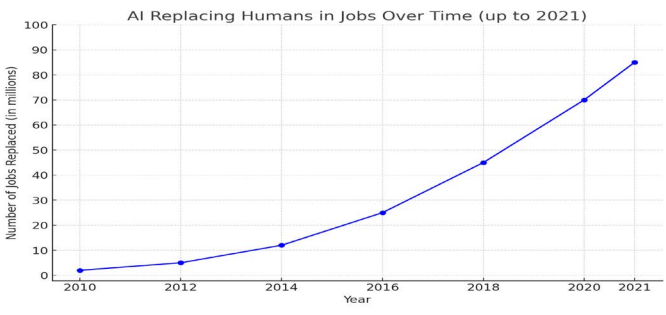

Figure 1: AI replacing human in jobs over time (up to 2021).
Source: The Future of Jobs Report 2020[3]

Traditionally, AI was with a rule-based system and logic to perform tasks. One of the best examples of the rule-baked system is the General Problem Solver (GPS), which was developed in the 1950s. This model aimed to solve human problems. However, this system was limited by its rigidity and inability to handle complex, real-world scenarios. The emergence of machine learning took place in the 1980s, and it is a subset of artificial intelligence in AI development. Unlike rule-based systems, machine learning algorithms could learn data from and improve their performance over time. This resulted in various machine learning models like k-nearest neighbors and support vector machines, which work for more sophisticated tools[7].

The evolution of AI has had a profound impact on several sectors. Here is a pie chart showing the impact

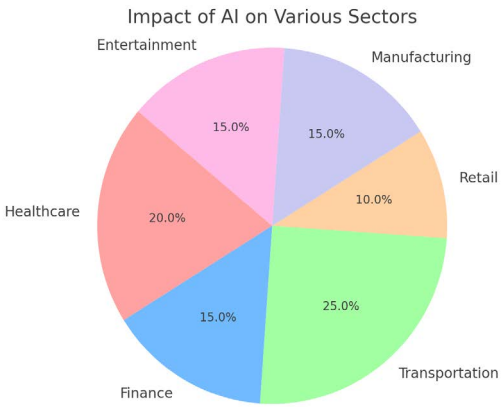

Figure 2: Impact of AI on various sectors.

Source: How artificial intelligence can make healthcare human again[2]

1. Ethical Issues in AI

Adoption of These technologies comes with a wide range of concerns that arise from the development, deployment, and utilization of artificial intelligence. The origin of ethical considerations in AI can be traced back to several vital factors, one of which is bias and discrimination because they have been trained on historical data. If the biases are not addressed, AI can promptly amplify discriminatory practices. For example, facial recognition has a high range of error rates on specific demographics, leading to complaints about fairness and equality[6].

Another is accountability and transparency. AI decisions are often opaque, making it difficult to determine how decisions are made. This lack of transparency poses a challenge to holding AI systems accountable for their actions. For Instnca, can AI make harmful medical diagnostic or unfair hiring decisions? It is crucial to determine who is responsible for the outcome.

These ethical issues can have different implications, such as social inequality, because AI can reinforce existing social equality, leading to unfair treatment of marginalized groups. This can exclude social injustice and equality, making it essential to develop AI systems that are fair and unbiased. The displacement of AI has significant implications, resulting in a rise in unemployment and social unrest. Tackling these challenges requires proactive policies to support the affected workers and promote equitable economic growth.

2. Biases in AI Systems

Bias mainly refers to the systematic and unfair discrimination of particular groups of people or individual attributes with AI algorithms and systems. Several reasons can cause biases in AI systems. This mainly comes with bias in training data because, first, the machine learning modes that are trained do not represent a diverse population or contain biased information. The design of an AI algorithm can indirectly introduce bias. This mainly depends upon the selection of features, the choice of model, or optimization criteria[1].

One of the prominent examples of bias in AI systems that took place in the 2016 investigation was COMPAS, a risk assessment tool used in the US criminal justice system. The study found that COMPAS was biased against African-American defendants, who were more likely to be incorrectly classified as high-risk compared to white defendants. Specifically, the tool falsely flagged black defendants as future criminals at nearly twice the rate of white defendants.

Another prominent example is Amazon's hiring algorithm, which was trained on historical data of 10 years. This machine learning model was utterly biased in the context of women, so they had to scrape the AI model[4].

A 2019 study found that an AI algorithm used to manage the healthcare population exhibited racial bias by allocating fewer resources to black patients compared to white patients with similar health conditions. The algorithm used for health needs inadvertently disadvantages black patients who historically incurred lower healthcare. The bias in healthcare algorithms can exacerbate health disparities, leading to inadequate care for minority patients and worsening health outcomes in already underserved communities[3].

3. Solutions to Mitigate AI Biases

Currently, there are efforts to reduce bias in Artificial Intelligence encompassing various strategies, focusing on data, algorithms, and processes that have a diverse representation of datasets because we make sure that training data is varied and representative of all demographic groups is crucial. This helps in minimizing the risk of biases that arise from over-representation or under-representation of certain groups. Making AI systems more transparent and explainable helps stakeholders understand how decisions are made. This can involve using interpolable models or creating explanations for complex model decisions.

Some proposed solutions can be inclusive design practices in the development of AI systems that can help identify and mitigate biases from the outset. This also includes a diverse perspective on teams responsible for data collection, algorithm design, and system evaluation. It ensures collaboration between industry and government to share best practices, tools, and research findings related to AI fairness and ethics[5].

Here is a pie chart summarizing all the solutions to mitigate risks

4. Bottom Line

The rapid integration of artificial intelligence into various sectors has brought about significant advancements, improving efficiency and accuracy. However, this progress is accompanied by ethical concerns and biases that need to be addressed. As evidenced by studies, such as the 2018 quantitative impact study on facial recognition errors, AI systems can perpetuate and amplify existing social biases, leading to unfair treatment of marginalized groups.

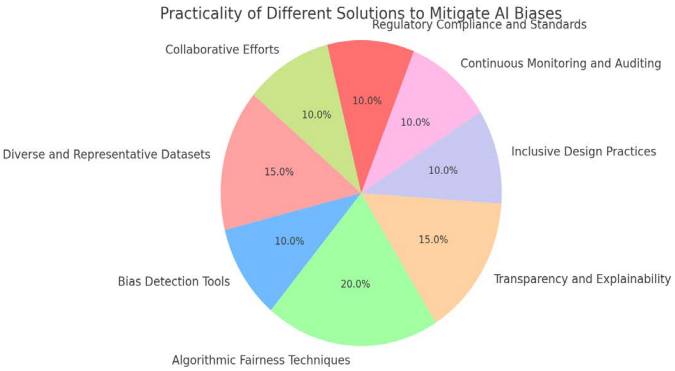

Figure 3: Practically of different solutions to migrate AI biases.

Source: Machine Bias: There's software used across the country[8]

The evolution from rule-based systems to advanced machine learning models has increased AI's capabilities but also intensified ethical issues. The origins of these biases often lie in biased training data, algorithmic design choices, and human biases. Real-world cases, such as the COMPAS risk assessment tool and Amazon's biased hiring algorithm, illustrate the severe impact of AI biases, affecting justice, employment, and healthcare outcomes.

Addressing these biases requires a multifaceted approach. Implementing diverse and representative datasets, using bias detection tools, and developing fairness-aware algorithms are crucial steps. Transparency and explainability in AI systems can help in understanding and mitigating biases. Moreover, inclusive design practices and continuous monitoring are essential to ensure fairness from the outset. Collaborative efforts among academia, industry, and government can promote best practices and ethical standards.

5. References

1. Angwin J, Larson J, Mattu S, Kirchner L. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica 2016.
2. Brynjolfsson E, McAfee A. The second machine age: Work, progress, and prosperity in a time of brilliant technologies. APA PsycNet 2014.
3. Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters 2018.
4. Noble SU. Algorithms of Oppression: How search engines reinforce racism. NYU Press 2018.
5. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366: 447-453.
6. Pasquale F. The black box society: The secret algorithms that control money and information. Harvard University Press 2015.
7. Simonite T. When it comes to gorillas, google photos remains blind. Wired 2018.
8. Zuboff S. The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs 2019.