Pranav Gokavarapu

Ms. Sarah Sowden

DATA 605

1 December 2023

Issue of algorithmic bias in Artificial Intelligence

INTRODUCTION

The development of artificial intelligence (AI) has revolutionized modern society and is now a vital factor behind technical breakthroughs in a wide range of fields. AI has become a key component of innovation, altering decision-making processes, and increasing productivity in difficult activities. However, algorithmic bias has emerged as a concern amid AI's remarkable development, which poses a significant barrier to the ethical application of AI. With its ability to evaluate vast amounts of data and provide insights that aid in decision-making, artificial intelligence has become an indispensable tool in the field of decision-making. This is seen in sectors like business and finance, where AI algorithms improve supply chain management and resource allocation and assist investment plans. However, using AI to make decisions raises questions about fairness, responsibility, and openness. AI has proven to have the ability to completely transform medication discovery, individualized treatment programs, and diagnostics in the healthcare industry. For example, AI-powered medical imaging improves diagnosis accuracy, but there is a chance that algorithmic bias in training data would result in unequal healthcare outcomes. Predictive policing algorithms in law and order also seek to maximize

resource allocation and deter crime. However, these instruments have faced criticism for maintaining prejudices found in past crime statistics, which can exacerbate already-existing social disparities. The ethical ramifications of algorithmic prejudice become more apparent as AI applications penetrate more areas of our lives. Promoting equitable chances and preventing discriminatory outcomes depend on ensuring fairness and reducing biases in AI systems. The rights and welfare of people in a variety of fields are directly impacted by the significance of ethics in AI, making it more than just a theoretical issue. We will explore the ethical implications of algorithmic bias in greater detail in this paper, using utilitarianism as our analytical framework. Through assessing the ramifications and overall influence on society of biased AI systems, we hope to highlight the necessity of creating moral standards and oversight frameworks. We will dissect algorithmic bias as we examine cases and examples, working to open the door for a more just and responsible application of AI in our quickly changing society. ("Teaching AI Ethics: Bias and Discrimination").

Case Briefing

One of the most pressing ethical issues in the field of artificial intelligence (AI) is known as "algorithmic bias," which is the phenomenon wherein the data used to train AI systems unintentionally reflects societal biases and prejudices. Because AI systems are trained on historical data and are inspired by human thought processes, they naturally incorporate the possibility of bias, which leads to discriminatory outputs. Although the goal is to develop algorithms that are unbiased and objective, the truth is that these systems are designed by people who have biases of their own, whether conscious or unconscious. The risk of sustaining and exacerbating human biases grows as we navigate an era where AI systems have a greater say over important aspects of our lives, like job opportunities and financial decisions. There are

several ways that algorithmic bias can appear in the context of AI. For instance, the AI system may unintentionally support discriminatory hiring practices if the historical data used to train algorithms reflects gender or racial biases in the hiring process. Similarly, there is a significant chance that law enforcement actions in predictive policing, where algorithms rely on past crime data, will reinforce preexisting prejudices. These instances highlight how important it is to examine and deal with algorithmic bias to maintain justice, fairness, and the avoidance of sustaining social injustices. (Anneroth, M. 2023; "Teaching AI Ethics: Bias and Discrimination").

Examples for bias and unfairness in AI applications

Police investigation. Predictive policing's introduction in the US has raised serious ethical questions about prejudice and discrimination in the criminal justice system. To predict crime hotspots and possible offenders, predictive policing makes use of data analysis, machine learning, and artificial intelligence. Critics contend that despite the goal of improving resource allocation and crime prevention, systemic inequalities may be sustained by inherent biases in the datasets used. For example, historical biases in law enforcement practices are reflected in the overrepresentation of Black people and people of color in police mugshot databases, which are used to train predictive algorithms. In August 2016, a coalition of 17 organizations, including the American Civil Liberties Union (ACLU), expressed concerns about the possibility of unfair and discriminatory outcomes. Additionally, the ethical considerations extend to facial recognition technology, which poses unique risks for historically marginalized communities due to systemic biases and inaccuracies, particularly against individuals with darker skin. The ethical discourse surrounding the appropriateness of using facial recognition for security purposes underscores the need for comprehensive guardrails to ensure equitable and just application of enhanced

surveillance technologies. (“Police Surveillance and Facial Recognition: Why Data Privacy Is Imperative for Communities of Color | Brookings”; “Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations | American Civil Liberties Union”).

Finance. Artificial Intelligence (AI) algorithms are fundamental to a variety of applications in finance and banking, from risk management for credit scoring, loan allocations, and mortgage rates to market forecasting for trading. There have been several cases where the decisions made based on these applications were considered unfair and biased, especially regarding borrowers who are minorities. Notably, research has shown that Black and Hispanic borrowers in the US have greater rates of loan and mortgage rejection, highlighting structural differences in financial resource availability. In addition, biases based on gender have been brought to light. For example, there have been cases where women have been given lower credit limits than men even though they have the same credit-relevant characteristics. These results highlight the necessity of critically analyzing and correcting biases in AI algorithms used in the financial industry. (Alejandra Bringas Colmenarejo et al.).

Health care. Artificial Intelligence (AI) technology is becoming more and more integrated in the healthcare industry, opening revolutionary possibilities such as AI-augmented clinical research and algorithms supporting image analysis and disease prediction. Artificial intelligence (AI) applications in surgery have demonstrated potential in predicting surgical outcomes, assisting surgeons with intraoperative navigation using computer vision, and evaluating technical proficiency and surgical performance. But algorithmic bias, a serious issue in the medical field, still poses a threat to the use of AI in surgery. This worry is illustrated by a noteworthy case by Kiyasseh et al., in which Surgical AI Systems (SAIS) were used to evaluate surgeon skill using

robotic surgery videos from various hospitals. Although the SAIS demonstrated consistency in assessing surgical performance, Biases in the form of over skilling and under skilling were exposed. The AI model incorrectly downgraded surgical performance, resulting in an under skilled outcome where a particular skill was predicted to be of lower quality than it was. On the other hand, over skilling happened when the AI model mistakenly improved surgical performance, believing that a particular skill was more proficient than it was. The existence of these biases, as demonstrated in this instance, highlights how critical it is to address algorithmic bias in medical AI applications to guarantee fair and accurate evaluations and, ultimately, protect the standard of patient care. (Mittermaier et al.).

ETHICAL THEORY: UTILITARIANISM

As a consequentialist ethical theory, utilitarianism emphasizes the greatest good for the greatest number of people when determining the morality of a given course of action. Utilitarianism, which is based on the maximization of well-being and minimization of harm, offers a framework for evaluating the moral implications of decisions made in the context of algorithmic bias and artificial intelligence (AI). The application of utilitarian principles becomes critical considering the growing integration of AI technologies across a range of sectors, including law enforcement, healthcare, and finance. According to this ethical viewpoint, the significance of a choice or course of action is determined by its capacity to benefit society, especially when it comes to AI systems that might unintentionally reinforce prejudices and have detrimental and discriminatory effects. To tackle algorithmic bias in these domains, a multimodal strategy that adheres to utilitarian principles is necessary. Minimizing biases is made possible by thoughtful data governance procedures, transparent AI algorithms, and the careful selection and curation of training data. In addition, prominent scientist Joy Buolamwini suggests a three-phase approach to

reduce bias: first, figuring out where bias exists in AI systems; second, inclusively curating datasets to address underrepresentation; and third, creating algorithms responsibly by taking ethical considerations into account. These steps make it possible to develop AI systems that prioritize justice, inclusivity, and positive societal impact in accordance with utilitarian principles. As a result, utilitarianism offers practical advice for developing morally sound and impartial AI systems that advance society in addition to serving as a theoretical framework for evaluating AI ethics. (TED; Joy Buolamwini).

CONCLUSION

In summary, the incorporation of Artificial Intelligence (AI) across diverse industries holds significant promise for advancing society; however, it is accompanied by the pressing issue of algorithmic bias. As a consequentialist ethical theory, utilitarianism serves as an essential prism through which to view the wider societal effects of AI systems that are biased. When developing and implementing AI technologies, the application of utilitarian principles highlights the significance of maximizing well-being, minimizing harm, and giving fairness top priority. Nonetheless, the difficulties posed by algorithmic bias demand an all-encompassing strategy.

A multifaceted approach is needed to address bias in AI, with meticulous data curation, algorithm transparency, and responsible governance practices acting as essential elements. This method gains useful depth from renowned researcher Joy Buolamwini's three-step solution, which focuses on identifying bias in AI systems and curating datasets. inclusively to lessen underrepresentation and thoughtfully when creating algorithms to take ethical considerations into account. Developers can design AI systems that actively promote societal well-being while adhering to utilitarian principles by putting these guidelines into practice.

The ethical imperative is still very much in place as AI technologies continue to shape our future: strong measures that reduce bias and preserve the values of justice and inclusivity must be incorporated. Because utilitarianism places a strong emphasis on the greater good, it emphasizes the necessity of continuing oversight, openness, and ethical considerations in the creation and application of AI. We can work to develop AI systems that actively contribute to a more just and positive social landscape rather than just avoiding bias perpetuation by following these guidelines and incorporating workable solutions. (TED; Joy Buolamwini).

Works Cited:

Mittermaier, Mirja, et al. “Bias in AI-based Models for Medical Applications: Challenges and Mitigation Strategies.” *Npj Digital Medicine*, vol. 6, no. 1, June 2023, doi:10.1038/s41746-023-00858-z.

Anneroth, M. (2023, September 4). AI bias and human rights: Why ethical AI matters. Telefonaktiebolaget LM Ericsson. https://www.ericsson.com/en/blog/2021/11/ai-bias-what-is-it

“Police Surveillance and Facial Recognition: Why Data Privacy Is Imperative for Communities of Color | Brookings.” *Brookings*, 27 Sept. 2023, www.brookings.edu/articles/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color.

“Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations | American Civil Liberties Union.” *American Civil Liberties Union*, 31 Aug. 2016, www.aclu.org/documents/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice.

Alejandra Bringas Colmenarejo1, Luca Nannini2, Alisa Rieger3, Kristen M. Scott4, XuanZhao5, Gourab K. Patro6, Gjergji Kasneci7, and Katharina Kinder-Kurlanda8 "Fairness in Agreement With European Values: An Interdisciplinary Perspective on AI Regulation" arXiv, 2207.01510, 2022, https://arxiv.org/pdf/2207.01510.pdf.

“Teaching AI Ethics: Bias and Discrimination.” *Leon Furze*, 7 Mar. 2023, leonfurze.com/2023/03/06/teaching-ai-ethics-bias-and-discrimination.

TED. “How I’m Fighting Bias in Algorithms | Joy Buolamwini.” *YouTube*, 29 Mar. 2017, www.youtube.com/watch?v=UG_X_7g63rY.