Chapter 3
Bias and Fairness in AI Technology

Muhsina
PES University, India

Zidan Kachhi
https://orcid.org/0000-0002-8317-6356
PES University, India

ABSTRACT

This chapter's objective is to provide an overview of how artificial intelligence (AI) has become an essential part of human life. It explains the sources of bias and its types in AI technology. With the help of previous studies, the chapter elucidates the strategies that can be used to avoid decision-making as a source of bias in AI technology. It also talks about the importance of understanding how human bias can also cause AI systems to exhibit bias towards certain groups. Unfairness in AI is also one of the most common sources of biassed data, and it's explained with strategies for detecting and addressing unfairness. The chapter also covers the need for transparency in AI technology along with ethical considerations, as transparency in AI is essential to ensuring that AI systems operate in adherence to ethical standards.

INTRODUCTION

Artificial intelligence (AI) has become an integral part of our lives and has transformed the way we interact with technology. With the advancement of AI, there has also been a growing concern about the potential for bias and unfairness in AI technology. In recent years, there have been several high-profile cases where AI systems have demonstrated bias and unfairness, which has led to a growing concern about the ethical implications of AI. This chapter will examine the issue of bias and fairness in AI technology, including its causes, effects, and potential solutions. The chapter will also explore the ethical implications of biased AI and provide recommendations for addressing this issue.

AI bias often comes from the data used in its training process. The factor that majorly affects AI performance is data quality. Like this, if an AI is trained with biased data, it will mirror that bias. Let's

DOI: 10.4018/979-8-3693-1565-1.ch003

say, if an AI is programmed mainly using data from a specific race or gender, it may struggle to correctly identify individuals of different races or genders. Lack of diversity in the groups responsible for AI development and training also adds to AI bias.

A team mainly having people from one type of background can lead to a biased AI system. When this happens, the bias can impact a lot of people and society as a whole. A big issue is when biased AI keeps old social inequalities going. For instance, an AI system might not fairly assess job candidates if it has a bias against women or certain racial or ethnic groups. This could lead to uneven job chances. Biased AI can also lead to unfair treatment of individuals. For example, an AI used for deciding if someone should be given credit might not be fair if it biases towards certain neighborhoods or income ranges. People could be refused credit or asked to pay more interest.

Bias in AI can be addressed in a number of ways. One strategy is to make sure that the AI systems' training data is representative of the whole population and diversified. To do this, gather data from multiple sources and make sure it is fair across demographics like gender and ethnicity. Diversifying the personnel in charge of creating and educating AI systems is another way to find a solution. A diverse team diminishes the likelihood of individual biases affecting policy creation. Last, uphold ethical benchmarks. Recently, a significant problem regarding AI fairness and equality has emerged, catching scholars, researchers, and decision-makers' attention. Studies have made progress in onset, outcome, and potential solutions.

Research from Buolamwini and Gebru (2018) revealed an interesting point. Major tech companies like IBM, Microsoft, and Face++ have developed facial recognition software. But they found a problem. The software shows more mistakes for women and people of color. Surprising, right? Their conclusion provides a clue. A lack of diversity in training data causes bias in the software. This means the software isn't as reliable as we'd expect!

The authors suggested companies making face recognition software need to have varied, representative training data to stop bias in their software. Mittelstadt et al. (2016) study, looked at how biased AI could impact ethics. They said biased AI might make social unfairness worse, even causing some people to face discrimination. The authors said that AI tech creators should include fair and unbiased aspects in their AI design and development stages.

Artificial Intelligence technology has advanced significantly, but it still displays prejudice and often falls short when it comes to fairness. This presents some significant challenges that demand urgent solutions. Firstly, it's crucial that we produce more diverse data sources for training AI systems. The lack of diversity in the current data often results in biased outcomes and increased inequality. Secondly, there's no standard methodology to detect and mitigate bias across all fields of AI. This hinders the creation of effective, versatile solutions. Lastly, in dealing with the complex social and technical problems of AI bias, interdisciplinary research that encompasses fields like ethics and sociology is needed more than ever. In fact, the everchanging character of AI systems makes it tough to consistently oversee and renew algorithms for lasting fairness. Likewise, little investigation exists on the lifelong effects of partial AI systems on society, such as its consequences on democracy, social unity, and human rights. It is crucial to bridge these gaps in research to promote ethical AI tech growth that fetches benefits for every person in society.

Another need to explore bias and fairness in AI is transparency and accountability. AI can often be puzzling, making it hard to see and fix biases. By digging into these topics, we can make sure AI is not only transparent and answerable but also impartial. Focusing on bias and fairness when creating AI helps make them more ethical and responsible. As AI gets a bigger role in our regular activities, we need to make sure it's suitable for everyone. This means designing AI that respects privacy, safeguards personal