

Chapter 5

Bias and Fairness

Addressing

Discrimination in

AI Systems

Padmaja Pulivarthy

Independent Researcher, USA

Pawan Whig

 <https://orcid.org/0000-0003-1863-1591>

VIPS, India

ABSTRACT

As artificial intelligence (AI) becomes increasingly pervasive in decision-making processes across various sectors, concerns about bias and fairness have risen to the forefront of ethical discussions. This chapter delves into the complex landscape of bias in AI systems, exploring its origins, manifestations, and implications for societal equity. We examine how biases can inadvertently infiltrate algorithms through data collection, preprocessing, and model training phases, leading to discriminatory outcomes against certain demographic groups. Moreover, we explore methodologies and frameworks aimed at mitigating bias, such as fairness-aware algorithms, bias detection techniques, and diversity-enhancing approaches. Ethical considerations and regulatory efforts are also scrutinized, highlighting the urgent need for transparency and accountability in AI development. By addressing these issues comprehensively, this chapter aims to contribute to the ongoing dialogue on fostering inclusive and equitable AI systems that uphold fundamental human rights and dignity.

DOI: 10.4018/979-8-3693-4147-6.ch005

Copyright © 2025, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

INTRODUCTION

In the realm of artificial intelligence (AI), the promise of enhancing efficiency, accuracy, and decision-making capabilities has propelled its integration into diverse sectors of society, from healthcare to finance, and from criminal justice to education. AI systems, driven by sophisticated algorithms and fueled by vast amounts of data, have the potential to revolutionize how we work, live, and interact with the world around us. However, amid these advancements lurks a profound challenge: the issue of bias and fairness in AI systems.

This introduction sets the stage by exploring the dual-edged sword of AI technology—its transformative potential and the ethical dilemmas it poses. We begin by defining bias in the context of AI and unpacking its multifaceted nature. Bias, in this context, refers to the systematic and unfair preferences or prejudices that AI systems may exhibit, leading to discriminatory outcomes that disadvantage certain individuals or groups based on factors such as race, gender, socioeconomic status, or other protected characteristics.

The Rise of AI and Its Ethical Imperatives

The rapid proliferation of AI technologies has outpaced the ethical frameworks needed to govern their development and deployment responsibly. As AI systems autonomously make decisions that impact people's lives—such as determining creditworthiness, predicting criminal recidivism, or selecting job candidates—the stakes for ensuring fairness and equity have never been higher. The decisions made by these systems can perpetuate existing inequalities or amplify societal biases, thereby undermining principles of fairness, justice, and human rights.

Understanding Bias in AI Systems

To grasp the intricacies of bias in AI, we delve into its underlying causes and manifestations. Bias can originate from multiple sources within the AI pipeline: from biased data used to train algorithms, to the design choices made in developing those algorithms, and even to the context in which AI systems are deployed. For instance, historical biases embedded in training data—reflecting societal inequalities and prejudices—can be inadvertently learned and perpetuated by AI models, reinforcing disparities in outcomes across different demographic groups.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/bias-and-fairness-addressing-discrimination-in-ai-systems/359640

Related Content

Generative AI's Impact on the Hospitality Industry

Anam Afaq, Meenu Chaudhary, Loveleen Gaurand Rajender Kumar (2025). *Generative Artificial Intelligence and Ethics: Standards, Guidelines, and Best Practices* (pp. 243-266).

www.irma-international.org/chapter/generative-ais-impact-on-the-hospitality-industry/358934

Sealing One's Online Wall Off From Outsiders: Determinants of the Use of Facebook's Privacy Settings Among Young Dutch Users

Ardion Beldad (2019). *Cyber Law, Privacy, and Security: Concepts, Methodologies, Tools, and Applications* (pp. 854-869).

www.irma-international.org/chapter/sealing-ones-online-wall-off-from-outsiders/228759

Empowering Ethical AI Adoption: Bridging the Gaps With Ethics4AI Platform

Ghinwa Jallouland Youssef M. Jalloul (2026). *The Ethical Landscape of AI: Global Issues and Solutions* (pp. 111-148).

www.irma-international.org/chapter/empowering-ethical-ai-adoption/399864

Ethical and Privacy Implications of the Use of Social Media During the Eyjafjallajökull Eruption Crisis

Hayley Watson and Rachel L. Finn (2019). *Cyber Law, Privacy, and Security: Concepts, Methodologies, Tools, and Applications* (pp. 764-777).

www.irma-international.org/chapter/ethical-and-privacy-implications-of-the-use-of-social-media-during-the-eyjafjallajokull-eruption-crisis/228754

Attacks on Web Applications

Ayushi Malik, Shagun Gehlot and Ambika Aggarwal (2023). *Perspectives on Ethical Hacking and Penetration Testing* (pp. 31-62).

www.irma-international.org/chapter/attacks-on-web-applications/330259