INTERROGATING ALGORITHMIC FAIRNESS: A PHILOSOPHICAL EXPLORATION OF JUSTICE AND BIAS IN MACHINE LEARNING

Etaoghene Paul Polo

Department of Philosophy, University of Delta, Agbor, Delta State, Nigeria

etaoghenepaulpolo@gmail.com

Abstract

As machine learning (ML) systems become increasingly embedded in areas such as healthcare, education, hiring, and criminal justice, concerns about fairness and bias have intensified. This paper explores what it means for an algorithm to be fair, focusing on the concept of justice and how it can guide the design and evaluation of ML systems. Drawing insights from social and political philosophy, particularly theories of distributive justice and equality of opportunity, the paper examines the strengths and limitations of common fairness metrics, such as demographic parity and equalised odds. These metrics, while useful, often fall short in addressing the broader ethical and societal challenges posed by biased algorithms. The paper argues for a deeper understanding of fairness that goes beyond technical solutions; one that considers the ethical, social, and systemic dimensions of justice. By connecting the bioethical principle of justice to fairness in ML, the paper proposes a more comprehensive framework to better navigate the complexities of fairness-aware learning, while ensuring greater accountability and transparency in AI systems. The aim is to bridge the gap between technical approaches to fairness and their real-world impacts, encouraging interdisciplinary collaboration to shape more equitable and responsible AI technologies.

Keywords: Algorithmic Fairness and Accountability, Machine Learning Bias, Systemic Bias in AI, Distributive Justice, Justice in AI Systems, Ethical AI, Responsible AI, AI Transparency, Accountability in AI Systems, Accountability in Machine Learning, Equality of Opportunity, Bioethics, Fairness Metrics, Social and Political Philosophy, Fairness-aware Learning