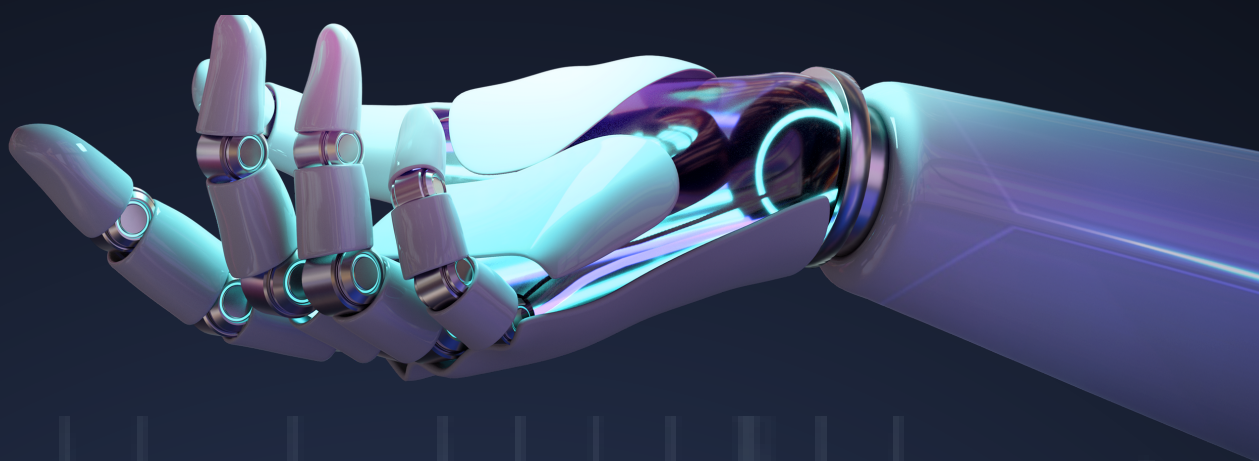


ARTIFICIAL INTELLIGENCE

GROUP FAIRNESS IN ML



INTRODUCTION

- Fairness in machine learning refers to the equitable treatment of individuals or groups, regardless of their demographic characteristics. It aims to mitigate biases and discrimination in algorithmic decision-making processes.

DEFINITION

- Disparate Impact: Measures whether algorithmic decisions disproportionately impact different groups. For example, a hiring algorithm that favors one demographic group over another may exhibit disparate impact.
- Equal Opportunity: Ensures that predictive models provide equal opportunities for positive outcomes across demographic groups. For instance, an admission model should predict student success equally well for all racial groups.
- Demographic Parity: Requires that the distribution of outcomes is consistent across demographic groups. For instance, a loan approval model should approve loans at similar rates for all racial groups.

COMPARISON

- Each group fairness metric has its strengths and limitations. Disparate impact is sensitive to imbalances in group representation but does not guarantee equal opportunities. Equal opportunity ensures fairness in outcomes but may ignore disparities in group representation. Demographic parity aims for proportional representation but may not address underlying biases.

IMPACTS

- Biased algorithmic decisions can have significant real-world impacts, perpetuating discrimination and exacerbating societal inequalities. For example, biased hiring algorithms can perpetuate gender or racial disparities in employment opportunities, leading to reduced diversity and economic inequality.

ENSURING FAIRNESS

- Collect diverse and representative data to mitigate biases in machine learning models.
- Regularly audit and evaluate models for fairness using appropriate group fairness metrics.
- Incorporate fairness considerations into the entire machine learning pipeline, from data collection to model deployment.
- Implement transparency and accountability mechanisms to ensure the responsible use of machine learning algorithms.