

Bias in AI



Bias Sources

AI inherits biases from training data, risking unfair discrimination.



Real-World Examples

Facial recognition often misidentifies minorities; hiring algorithms favor certain groups.



Mitigation Methods

Techniques include data rebalancing and adversarial debiasing to reduce bias.



Persistent Challenges

Hidden biases in complex AI models remain difficult to fully measure or eliminate.

Fairness in AI

Fairness ensures AI decisions do not disadvantage groups. This requires inclusive datasets and equitable algorithm design.



Demographic Parity

Ensures outcomes are independent of sensitive attributes like race or gender.



Equal Opportunity

Focuses on equal true positive rates across groups to prevent discrimination.



Social Justice

Promotes non-discriminatory outcomes consistent with UNESCO's ethical principles.