Bias and Fairness

Dataset Bias

Selection bias, label bias. Underrepresentation of certain demographics

Demographic Disparities

Performance gaps across age, sex, race. Different disease prevalence and presentation

Fairness Metrics

Equalized odds, demographic parity. Group fairness vs individual fairness

Mitigation Strategies

Diverse training data, reweighting, adversarial debiasing. Fairness constraints

Continuous Monitoring

Track performance by subgroup. Detect and address emerging disparities