

Fairness Interventions Across the ML Pipeline

Fairness Metrics

Demographic Parity

Equal Opportunity

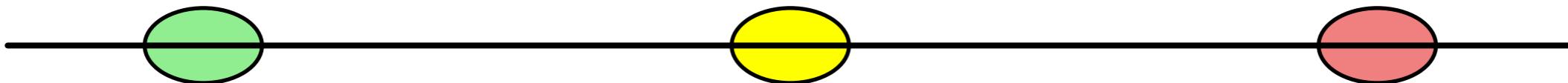
Equalized Odds

Individual Fairness

Pre-processing

In-processing

Post-processing



- Synthetic data
- Re-sampling
- Data augmentation

- Constraints
- Adversarial debiasing
- Fair objectives

- Fair ranking
- Threshold optimization
- Output calibration

Note: Different fairness metrics often conflict - perfect fairness is mathematically impossible