

# Group Fairness in Machine Learning

Group fairness in machine learning means making sure that algorithms treat all groups of people fairly, no matter their race, gender, or age. It aims to give equal results for everyone by finding and fixing any biases.

This fairness is based on three important ideas: independence, sufficiency, and separation.

## INDEPENDENCE

It ensures predictions are impartial to sensitive attributes, equalizing classification likelihood across diverse individuals.

### Demographic Parity

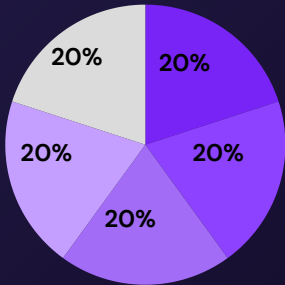
Equal distribution of favorable outcomes among sensitive / protected groups.

### Equal Selection Parity

Same number of favorable outcomes in sensitive / protected groups.

### Conditional Statistical Parity

Equal chances of favorable outcomes in sensitive/protected groups, accounting for legitimate attributes.



## SUFFICIENCY

It ensures that sensitive attributes do not influence the true outcome, guaranteeing fair decisions across all groups and individuals, regardless of their specific attributes or predictions.

### Conditional Use Accuracy Equality

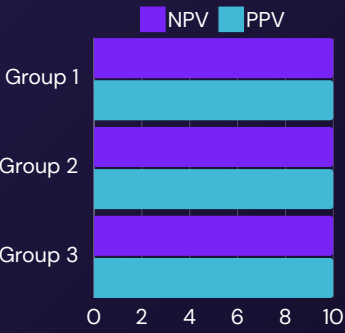
Uniform PPV and NPV across sensitive / protected groups.

### Predictive Parity

Consistent PPV among sensitive / protected groups.

### Calibration

Balanced PPV and NPV for all sensitive / protected groups by adjusting false discovery and omission rates.



## SEPARATION

It ensures that sensitive attributes do not affect predicted outcomes, guaranteeing fair predictions for all groups and individuals regardless of their unique attributes.

### Equalized Odds

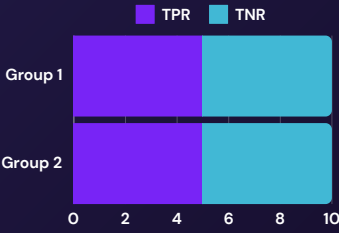
Uniform TPR and TNR across sensitive / protected groups.

### Equalized Opportunities

Consistent TPR across sensitive/protected groups.

### Predictive Equality

Equal TNR among sensitive/protected groups.



## USAGE EXAMPLE

1. **Employment:** Bias reduction in hiring algorithms for equal opportunities.
2. **Finance:** Equitable credit scoring, avoiding ethnic or socioeconomic biases.
3. **Education:** Fair educational algorithms for unbiased student evaluation.
4. **Housing:** Equality-focused housing algorithms to prevent race or gender discrimination.

## REAL-WORLD IMPACTS

1. **Cultural Sensitivity:** Fair algorithms respect and reflect cultural diversity.
2. **Accessibility:** Fair algorithms improve service access for diverse groups.
3. **Economic Efficiency:** Fairness in algorithms enhances resource allocation.
4. **Public Trust:** Unbiased algorithms boost confidence in technology.

## STATE-OF-THE-ART APPLICATIONS

Explainable AI (XAI) platforms, using group fairness metrics, enhance transparency in AI decision-making. This is pivotal in sectors like finance, healthcare, and criminal justice for bias detection and trust-building.

## BEST PRACTICES

1. **Data Representation:** Using diverse and representative data.
2. **Transparency:** Ensuring algorithmic transparency.
3. **Monitoring & Correction:** Continuous monitoring and implementing corrective actions.
4. **Community Involvement:** Engaging stakeholders and communities.
5. **Ethical Standards:** Adhering to ethical guidelines.
6. **Privacy & Security:** Implementing robust privacy and security measures.
7. **Training & Education:** Providing thorough training to users and decision-makers.
8. **Algorithm Maintenance:** Regularly updating and refining algorithms.
9. **Collaboration:** Encouraging interdisciplinary collaboration.
10. **Accountability:** Establishing clear accountability and governance structures.