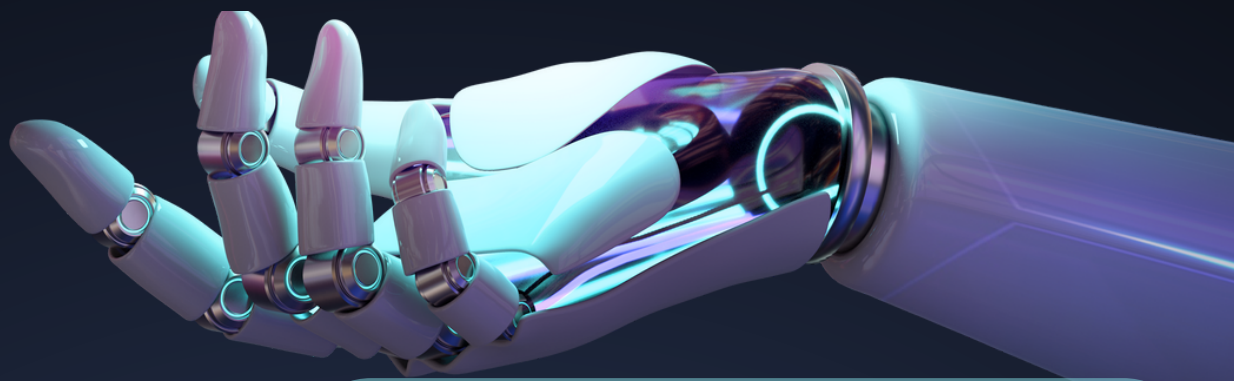


ARTIFICIAL INTELLIGENCE

ALL ABOUT GROUP FAIRNESS



WHAT IS IT?

Also referred to as statistical parity, **group fairness** is a requirement in the field of Artificial Intelligence (specifically machine learning) which states that protected groups should be treated similarly to the advantaged group or the populations as a whole.

The main difference between *group* fairness and *individual* fairness is that group fairness does not consider the individual merits and may result in choosing the less qualified members of a group, whereas individual fairness assumes a similarity metric of the individuals for the classification task at hand that is generally hard to find.

GROUP FAIRNESS METRICS

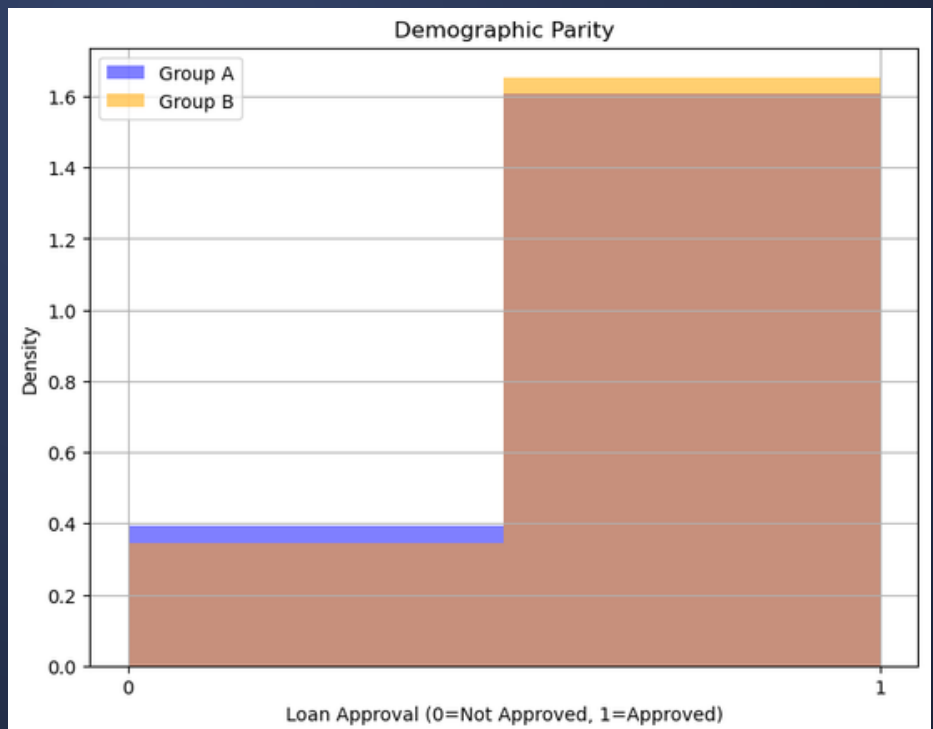
DEMOGRAPHIC PARITY

This metric checks whether the probability of a positive prediction is equal across all demographic groups.

Demographic parity is satisfied when their base rates are the same.

Proportion of predicted positives must be equal across sensitive/ protected groups.

EXAMPLE: an employer who wishes to have similar job acceptance rates for both male and female candidates. This means that 50% of male and 50% of female candidates get the job.

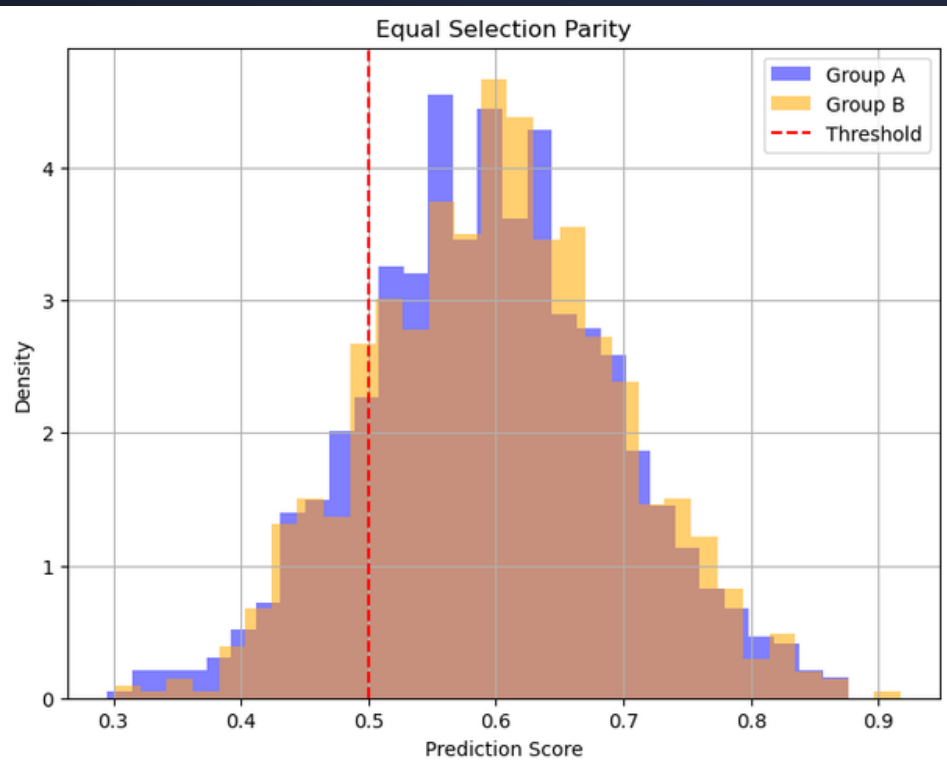


EQUAL SELECTION PARITY

This metric evaluates whether the proportion of selected individuals from each group is approximately the same. Equal selection parity is satisfied when the search engine returns the same number of images for each sensitive/protected attribute group.

Number of predicted positives must be equal across sensitive/ protected groups.

EXAMPLE: Suppose a company is using an algorithm to screen job applicants for interview invitations. The algorithm predicts the likelihood of an applicant's success based on various features such as education, work experience, and skills.



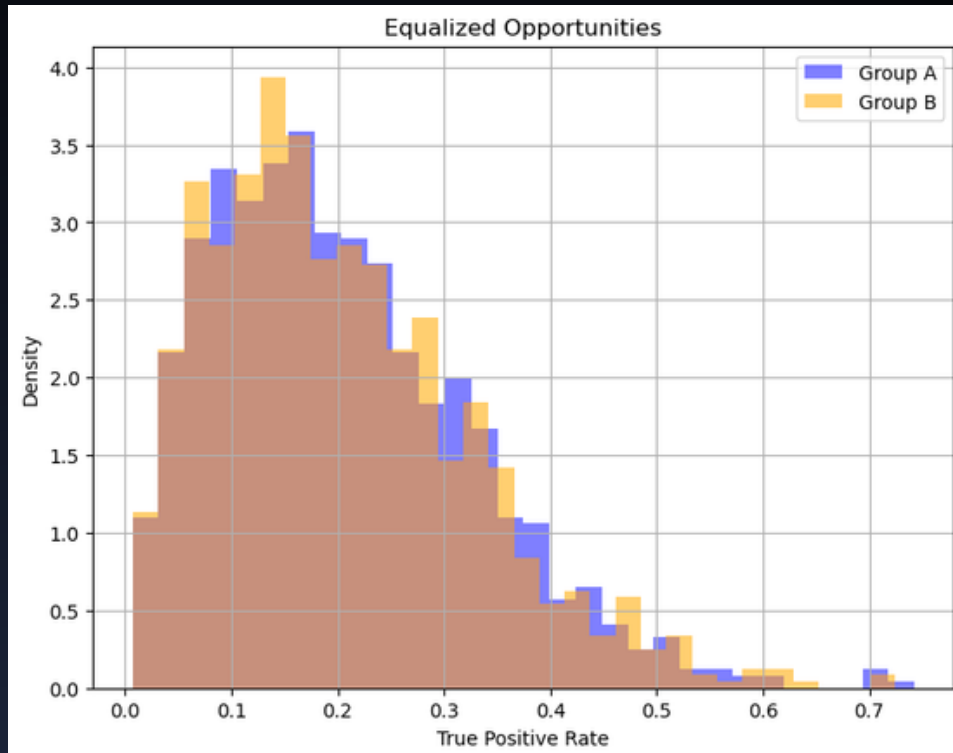
EQUALIZED OPPORTUNITIES

This metric measures whether the true positive rate is equal for all groups. It focuses on minimising false negatives.

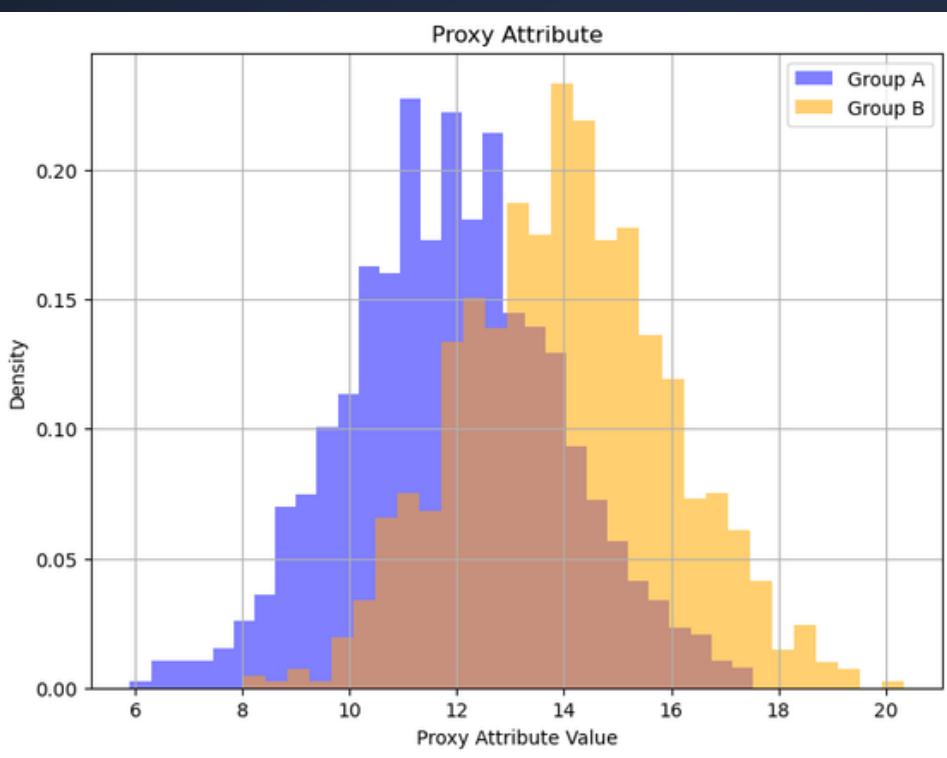
Equalized opportunities is satisfied when the return rates for an image label are equal across sensitive/protected groups.

The true positive rates (TPR) should be equal across sensitive/protective groups.

EXAMPLE: In loan approvals, a bank's machine learning algorithm ensures equal opportunity by training on diverse datasets and focusing solely on creditworthiness factors like income and credit history.



PROXY ATTRIBUTE



This metric is an observable variable that is used as a substitute or approximation for a protected attribute, which may be legally or ethically sensitive.

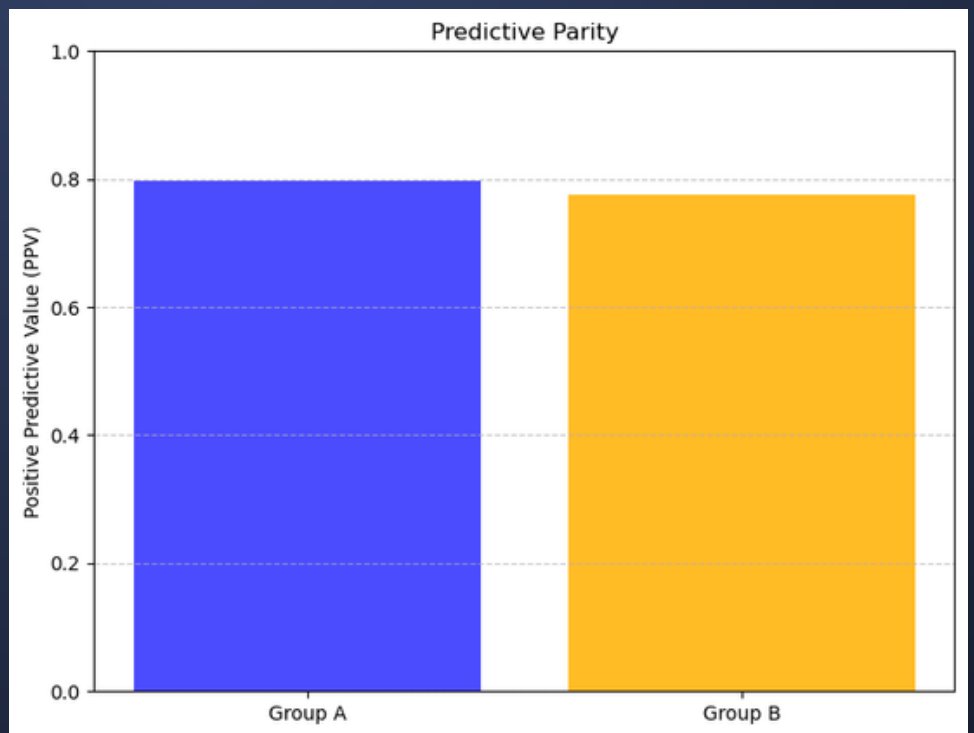
EXAMPLE: In a hiring process, educational attainment or zip code might be used as proxy attributes for race or socioeconomic status.

PREDICTIVE PARITY

This metric is used to assess the fairness of machine learning models, particularly in binary classification tasks.

Predictive parity/Precision/PPV is satisfied when the proportion of correctly returned images of a label is equal across sensitive/protected groups.

'The positive predictive value (PPV) should be equal across sensitive/protective groups'.



REFERENCES

Fairness measures. Fairness Measures - Detecting Algorithmic Discrimination. (n.d.). [<https://fairnessmeasures.github.io/Pages/Definitions>]

Gomede, E. (2023, September 5). Fairness Metrics in Machine Learning. Medium. [<https://medium.com/@evertongomede/fairness-metrics-in-machine-learning-8c3777b48a9c>]

Group fairness vs. individual fairness in machine learning. Lumenova AI. (2022, August 31). [<https://www.lumenova.ai/blog/group-fairness-vs-individual-fairness/>]

Zohar, Y. (n.d.). Fairness Metrics In Machine Learning. aporia. [<https://www.aporia.com/learn/fairness-metrics-in-machine-learning/>]

All the images have been generated by me. The code for them can be found here: [<https://github.com/BredaUniversityADSAI/2023-24c-fai1-adsai-dariavladutu236578/blob/main/Week%201%20-%20Responsible%20AI/Images%20for%20infographic.ipynb>]