

## User Manual: Fairness Testing Tool

This tool analyzes bias in machine learning models and applies fairness evaluation and mitigation strategies.

### Access the Tool

Use the Colab Notebook here:

[https://colab.research.google.com/drive/1nj5d2ITG\\_LCqkELe10p3dJxe7j4YnzqW?usp=sharing](https://colab.research.google.com/drive/1nj5d2ITG_LCqkELe10p3dJxe7j4YnzqW?usp=sharing)

### Features

- Train a classification model on the Adult Income dataset
- Evaluate group fairness using MetricFrame
- Perform targeted perturbation by flipping sensitive attributes
- Conduct intersectional fairness analysis (race + gender)
- Run individual disparate impact (IDI) testing
- Apply in-processing fairness mitigation using ExponentiatedGradient

### How to Run

1. Open the Colab notebook
2. Run all cells from top to bottom
3. Review outputs, charts, and flagged biased groups