

Fair vs. Unfair Outcomes Comparison

Provide two versions of a model output (fair and biased).

Identify what changed and why it is more ethical.

(Focus: investigation, small experiments, structured thinking)

Here is a clear, simple, and *viva-friendly* answer for the **Fair vs. Unfair Outcomes Comparison** activity:

Fair vs. Unfair Outcomes Comparison

Scenario:

A loan approval model predicts whether a person should get a loan.

Unfair (Biased) Model Output

Input:

- Income: ₹40,000/month
- Credit Score: 750
- Past Loan Repayments: Good
- **Gender: Female**

Biased Model Prediction: ❌ *Loan Rejected*

Reason (unethical):

The model incorrectly uses **gender** as a factor. Even though the applicant meets financial criteria, the model rejects the loan due to bias in the training data.

Fair (Ethical) Model Output

Input:

- Income: ₹40,000/month
- Credit Score: 750

- Past Loan Repayments: Good
- Gender: Female (ignored as it should be)

Fair Model Prediction: ✔ *Loan Approved*

Reason (ethical):

The model only uses **relevant financial features** such as income and credit score. Sensitive attributes like gender, religion, caste, or age are **removed** so the decision is based purely on financial ability.

What Changed & Why It Is More Ethical

Aspect	Unfair Model	Fair Model
Use of Sensitive Data	Uses gender → creates discrimination	Removes or ignores gender
Decision Basis	Influenced by bias in historical data	Based only on relevant features
Outcome	Equal applicants get different results	Equal applicants get equal treatment
Ethical Standard	Violates fairness & equality	Promotes unbiased and transparent decisions

Why It Is More Ethical:

- Protects individuals from discrimination.
- Ensures equal opportunity for all applicants.
- Builds trust in AI systems.
- Follows fairness guidelines like **non-discrimination, transparency, and accountability.**

If you want, I can also turn this into a **diagram, presentation slide, or short exam-ready paragraph.**



