

Lesson 47 Summary: ML Ethics
Data Science with Python – Key Concepts

Data Science Program

ML Ethics & Responsible AI



Responsible AI is essential for trust

Multiple definitions:

- **Demographic parity:** Equal acceptance rates
- **Equal opportunity:** Equal TPR across groups
- **Calibration:** Predictions match reality

No single fairness definition fits all contexts

Bias enters everywhere:

- **Data:** Historical discrimination
- **Features:** Proxy variables
- **Labels:** Human annotator bias

Awareness is the first step to mitigation

Transparency matters:

- **SHAP:** Feature importance per prediction
- **LIME:** Local interpretable explanations
- **Model cards:** Document model behavior

If you cannot explain it, you cannot trust it

High-stakes domain:

- **Credit:** Fair lending laws
- **Trading:** Market manipulation concerns
- **Insurance:** Discriminatory pricing

Financial decisions impact lives

Ethics Checklist:

Principle	Action
Fairness	Test across demographic groups
Transparency	Document model decisions
Privacy	Anonymize sensitive data
Accountability	Maintain audit trails
Human oversight	Keep humans in the loop

Ethics is a continuous process, not a checkbox