

AI Ethics Casefiles

Bias · Transparency · Responsibility

Why AI Ethics?

AI systems shape decisions in hiring, lending, healthcare, and justice. When they fail, consequences are not just technical — they're human.

Project Goals

- Investigate real-world AI ethics cases
 - Explore algorithmic bias and transparency
 - Demonstrate responsible design through code and critique
 - Communicate insights in accessible, visual formats
-

Case 1: Gender Bias in Hiring Algorithm

- System trained on historical hiring data
- Favored male candidates due to biased patterns
- Gender not explicitly used — but inferred via proxies

Ethical Issues:

- Proxy bias
 - Lack of explainability
 - Disparate impact
-

Case 2: Medical Risk Prediction


- AI model for patient risk assessment
- Underpredicted risk for Black patients
- Training data reflected systemic inequalities

Ethical Issues:

- Historical bias
 - Lack of fairness audits
 - No community engagement
-

Demo: Simulating Bias in a Classifier

- Toy dataset with gender and score
- Logistic regression model
- Gender feature increases likelihood of acceptance
- Highlights how bias can emerge from “neutral” data

 See `src/bias-demo.ipynb` for code

Transparency: Can We Explain These Systems?

- Most systems lack explainability
- Black-box models common in practice
- Efforts like LIME, SHAP, and model cards help — but are not enough

 Transparency must go beyond technical interpretability.

Responsible Design in Practice

- Audit datasets for imbalance
- Avoid using protected attributes — but watch for proxies
- Communicate model limitations clearly
- Engage affected communities

 **Tools:** Fairlearn, Aequitas, Model Cards, Datasheets

Limitations of This Work

- Educational scope, not a real audit
- Toy datasets oversimplify reality
- Focuses on specific cases; many dimensions of fairness not addressed
- No engagement with stakeholders or legal frameworks

Key Takeaways

- Bias is systemic, not just technical
- Transparency is ethical, not optional
- Design choices affect real lives
- Small demos can teach big lessons

Want to Learn More?

- FairMLBook.org
- [Aequitas](#)
- [LIME](#)
- [Model Cards](#)

Thank you!

Explore the full project:

github.com/Augusto047/ai-ethics-casefiles