

Ethical Reflection – Ensuring Ethical AI in My Projects

One of my recent projects involved auditing the COMPAS recidivism dataset using IBM's AI Fairness 360 toolkit to detect racial bias in risk scoring. This project made me reflect deeply on the ethical responsibilities of building and deploying AI systems that impact human lives.

In future AI projects — especially those involving sensitive domains like criminal justice, healthcare, or employment — I will take deliberate steps to ensure ethical AI principles are upheld.

First, I will prioritize **fairness** by actively testing for and mitigating bias during data collection, model training, and evaluation. This includes using tools like AIF360 or Fairlearn to detect disparities across protected groups such as race, gender, and age.

Second, I will ensure **transparency and explainability** by choosing interpretable models when possible and documenting how decisions are made. Users and stakeholders deserve to understand why an AI system produces a particular outcome.

Third, I will emphasize **accountability** by setting clear boundaries for how my models are used and regularly monitoring performance. Any system deployed must include mechanisms for human oversight, feedback, and redress.

Fourth, I will be vigilant about **privacy and data security**, especially when handling personal or sensitive information. This means complying with data protection laws and anonymizing data when appropriate.

Finally, I will engage in **inclusive design** by considering the needs and impacts of all users, especially those from marginalized communities.

By embedding these ethical principles into my development workflow, I aim to create AI solutions that are not only effective but also fair, responsible, and aligned with human values.

