

COMPAS Dataset Training - Ethics and Bias in Artificial Intelligence

Name: Joshua Ayad

Class: Introduction to Artificial Intelligence, Computer Science
311

Date: Monday, December 1st, 2025

The COMPAS dataset identifies profound ethical challenges between artificial intelligence and the field of criminal justice. COMPAS is a risk assessment tool designed to predict how likely someone is to be a repeat offender of the law. Use of COMPAS in courts has brought significant concerns to light regarding racial bias and fairness (Angwin et al., 2016). The dataset contains systemic disparities, especially between racially black and white defendants. For example, individuals identified as African-American are disproportionately labeled as "high risk," regardless of whether documents indicate that they reoffend. This reflects historical bias in criminal justice data, where over-policing and sentencing disparities skew underlying records.

When viewed through a model evaluation standpoint, relying solely on accuracy can obscure fairness concerns. The methods I

should be examined with a scrutinous eye when using any sort of system where bias has the nutrients to grow and multiply.

To responsibly use such models, developers and stakeholders must prioritize transparency, explainability, and fairness. Techniques like equal opportunity auditing, subgroup fairness metrics, and algorithmic bias minimization should be incorporated. AI systems should complement and not replace human judgment, particularly in sensitive domains like criminal justice. After all, the field of criminal justice is not a STEM field, and there are proper places to apply our vast knowledge of math and science, and there are places where we should not apply those tools.

This project offered valuable insights into the practical implementation of machine learning, as well as the ethical complexities that arise when applying AI to non-synthetic data. One of the first design decisions was selecting a dataset that would meaningfully allow for technical evaluation and fairness analysis. For this data, I selected the COMPAS dataset, because it is both widely used in criminal justice and controversial for its embedded racial bias. This made it ideal for exploring how algorithmic decisions can impact people's lives. A true scientist seeks the truth, and I felt that this dataset would lead us there.

Reflection on Design and Challenges

This project offered valuable insight into the practical implementations of machine learning, as well as the ethical complexities that arise when applying AI to non-synthetic, “real” data. One of the first design decisions involved selecting a dataset that would meaningfully allow for technical evaluation and fairness analysis.

At first, when I skimmed the assignment prompt, I wanted to use the NIST hand-written digit dataset, as, long before enrolling in this class I was working on a project that used that dataset. I also considered using OpenCV, which is an extremely popular and widely-used computer vision library for object recognition, since I have a few schematics and ideas in my personal knowledge base application related to that tool. However, as I read the rubric and assignment prompt further, I realized this wouldn’t aptly satisfy the “ethics” related goals of the project. There is a universe where facial recognition, especially in China, and their use of it for crime reduction and social credit, was my submission for this project. I also considered the use of facial recognition in sparsely rolled out police cameras stationed atop traffic lights and such, as well

as the use of facial recognition and data from those sources used by the corporation “Palantir”.

In the end, the COMPAS dataset was a compelling choice because it is both widely used in criminal justice and controversial for its embedded racial bias. This made it ideal for exploring how algorithmic decisions can impact people's lives in a more “emotional” and “ethical” way, with less subjectivity in the definition of “ethics”, since we often do not know what governments are really doing with these systems as would be the case with China or Palantir (based in the United States).

References (APA Style)

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016).

Machine bias. ProPublica.

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). *Algorithmic decision making and the cost of fairness*.

Proceedings of the 23rd ACM SIGKDD International Conference on

