

AI Tools Assignment – Ethical Reflection on Bias and Fairness

Name: Abba Imam

Course: AI Tools and Applications

Theme: Mastering the AI Toolkit

1. Introduction

AI systems are only as fair as the data and algorithms that power them. In our project, we explored three AI domains — classical ML (Scikit-learn), deep learning (TensorFlow), and NLP (spaCy). Each tool, while powerful, introduces ethical considerations in how data is processed and predictions are made.

2. Identified Biases

a. MNIST Digit Classifier (TensorFlow CNN):

Although MNIST appears neutral, it's limited to grayscale digits from a narrow dataset. This restricts model generalization to diverse handwriting or regional scripts. A model trained only on MNIST may perform poorly when deployed globally.

b. Amazon Product Reviews (spaCy):

The language used in reviews can reflect consumer sentiment biases — for example, negative bias toward specific product origins or brands. Additionally, imbalance between positive and negative reviews can distort sentiment models.

3. Mitigation Strategies

Issue	Tool-Based Mitigation	Explanation
Dataset imbalance	TensorFlow Fairness Indicators	Evaluate and visualize model performance across demographic slices to detect bias.
Rule-based bias	spaCy's Pattern Matching	Design deterministic patterns to avoid misclassification in culturally sensitive entities.
Overfitting to small samples	Data Augmentation & Stratified Sampling	Ensure fair representation of classes.

4. Ethical AI Principles

We aligned our workflow with key ethical principles:

- **Transparency:** We documented model architecture and data preprocessing steps.
 - **Accountability:** All models were peer-reviewed within our group.
 - **Fairness:** Bias checks were integrated into our validation process.
 - **Privacy:** No personally identifiable data was used in any experiment.
-

5. Reflection

Ethical AI is not an afterthought—it must be built into the design. Through TensorFlow and spaCy, we learned that fairness tools, explainability, and data diversity are essential for trustworthy AI. The challenge is continuous, but awareness and responsible engineering can bridge the gap between performance and ethics.

End of Report