Detecting AI-Generated Text — Proposal

Capstone: The Art of Approximation

Ojas Chaturvedi, Ritwik Jayaraman, Saianshul Vishnubhaktula, Zaheen Jamil

November 28, 2023

1. **Language:** Python, a simple and popular language for machine learning and data science due to its extensive libraries and frameworks
2. **Objective:** To develop a custom machine learning model which would be able to detect AI-generated text
3. **Implementation:**
   (a) **Overview of steps:**
      i. Data Exploration
      ii. Preprocessing
      iii. Feature Engineering
      iv. Model Development
      v. Training and Testing
      vi. Evaluation
      vii. Optimization
      viii. Documentation
   (b) **Potential Libraries:**
      i. **Pandas:** For data manipulation and analysis
      ii. **Matplotlib:** For visualizations
      iii. **NumPy:** For numerical computing and working with arrays
      iv. **Scikit-learn:** For data mining and analysis
      v. **TensorFlow:** For complex neural network modeling
      vi. **PyTorch:** For natural language processing
      vii. **NLTK/spaCy:** For human language data with symptom inputs
      viii. **Flask/Django:** For backend web development
      ix. **SQLAlchemy:** For SQL databases and Object-Relational Mapping
   (c) **Manual Work:**
      i. Compiling datasets
      ii. Building custom model
      iii. Training and testing model
      iv. Creating website/app
      v. Documentation of all steps
4. **Jobs:**
   (a) **Machine Learning Developers**
      i. Develops the machine learning model
      ii. Trains & Tests the model
      iii. Makes the model usable in the website/app
   (b) **Data Analyst**
      i. Data-exploration, preprocessing, and feature engineering
      ii. Will still contribute as a Machine Learning Developer
   (c) **GUI Developer**
      i. Makes the website/app and all of its functionality (UI)
      ii. Will still contribute as a Machine Learning Developer