***Title : Detecting machine generated text.***

***Group 23***

***Group Members: Lesedi Ntsele, Shiza Butt, Vutomi Mohube***

**Research Question**

The rapid rise of large language models like GPT has made it increasingly difficult to distinguish human-written texts from machine generated ones, especially in low-resource languages such as those spoken across Africa. This project aims to address the problem: *Can we develop robust models to accurately detect machine-generated texts in African languages, and what linguistic patterns differentiate them from human texts?* This is interesting because African languages are underrepresented in natural language processing research, and the proliferation of machine-generated content risks undermining trust in digital communication, especially in contexts like news or education where authenticity matters. Current detection methods are largely tailored to high-resource languages (e.g., English), leaving a gap in solutions for African languages that this project seeks to fill.

**Dataset and Data Description**

To effectively train and evaluate models for detecting machine-generated text in African languages, a balanced dataset comprising both human-written and machine-generated content will be used. Given the low-resource context of African languages, preprocessing it will ensure both quality and representativeness.The dataset will consist of both human-written and machine-generated text:

* **Common Crawl**: 10,000+ publicly available web text samples.
* **AfriSenti**: ~50,000 sentiment-labeled samples in African languages, available on GitHub.
* **Vukuzenzele**: ~5,000 public samples from a South African government source.
* **ChatGPT Data**: Synthetic data to be generated, matching the combined size of the human datasets (~65,000 samples).

### ***Attributes***

Each sample will include standardized attributes:

* text: the main content (sentence, paragraph, or article)
* source: data origin ( Common Crawl, AfriSenti, Vukuzenzele.)
* label: sentiment or topic label (where applicable)
* language: language of the sample
* type: data origin type (human\_text or synthetic)
* timestamp: date of publication or crawling
* source\_url: original URL

***Resources***

* ***Pretrained models -*** AfriBERTa
* ***NLP toolkits -*** NLTK

**Approach and Methodology**

In this model, classifiers that identify machine-generated material in African languages will be trained using a carefully selected dataset of human-written texts (Common Crawl, AfriSenti, Vukuzenzele) and texts produced by GPT. Tone and stringing of words are examples of language characteristics that will be addressed during preprocessing. To capture surface-level and contextual differences, statistical characteristics and embeddings from multilingual models such as XLM-RoBERTa and AfroLM will be employed. Because they blend interpretability and performance, traditional machine learning classifiers and fine-tuned transformers are used.Transfer learning will increase accuracy in low-resource languages, and explainability techniques (like SHAP and LIME) would guarantee clear, reliable predictions and offer insight into model behavior.

**Evaluation Strategy**

Success will be measured using standard classification metrics: accuracy, precision, recall, and F1-score, with a focus on F1 to balance false positives and negatives. We will perform 5-fold cross-validation to ensure robustness. Baselines include: (1) a random classifier (50% accuracy), (2) existing English-focused detection tools (e.g GLTR), and (3) a simple n-gram model. These benchmarks will help quantify our models’ improvement over naive or non-specialized approaches. Linguistic analysis will be evaluated qualitatively by comparing identified patterns against expert linguistic insights.

**Expected Outputs and Contributions**

A machine learning model that can identify machine-generated texts in African languages will be produced by this study. Particularly for underrepresented languages, it will offer direct insight into the linguistic characteristics of texts produced by AI as opposed to those created by humans. The research tackles the problem of data scarcity in low-resource environments by utilizing transfer learning with models such as XLM-RoBERTa and AfroLM. Explainability tools will increase transparency and facilitate the interpretation of model decisions. All things considered, the project advances ethical AI, enhances detection methods for multilingual situations, and promotes the digital inclusion of African languages in AI research, all of which advance the field of natural language processing.