Detecting Student vs. Large Language Model (LLM) Essays

TEAM MEMBERS:

ASHISH AGARWAL

Detecting whether an essay was written by a student or a Large Language Model (LLM)

Objective:

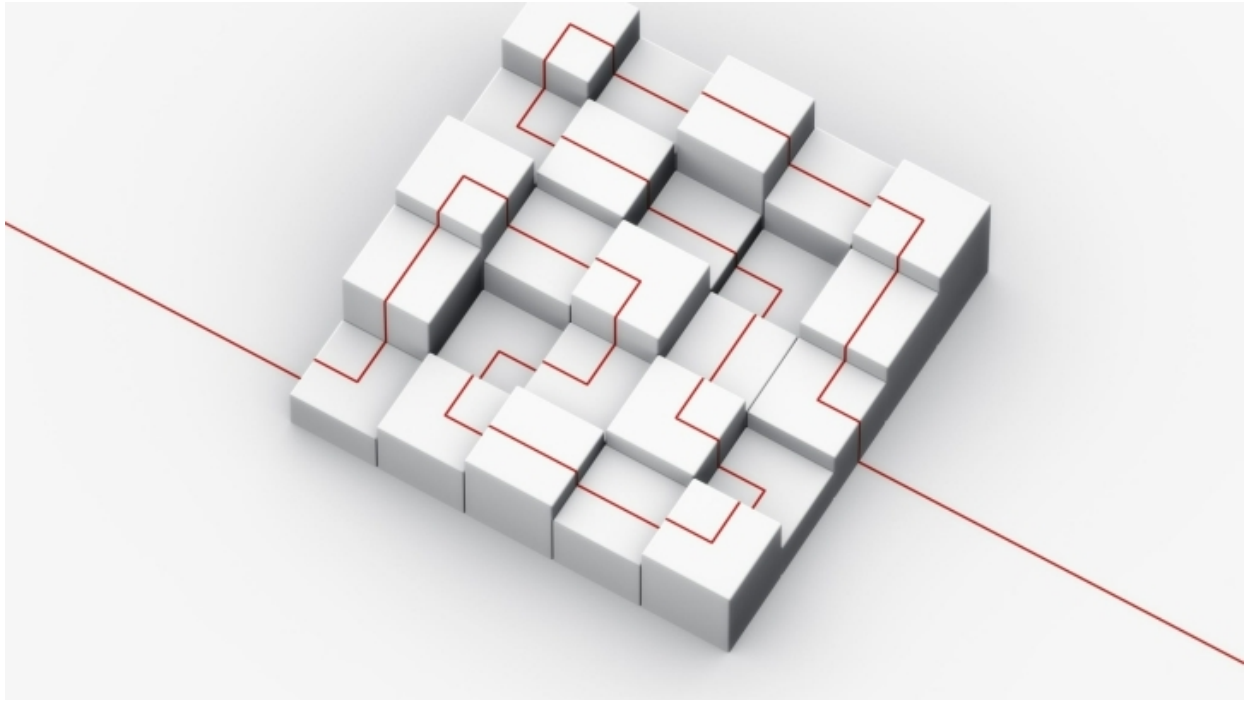
To develop a machine learning model capable of accurately identifying essays authored by middle and high school students versus those generated by LLMs.

Project Objectives:

- To build a robust machine learning model for distinguishing between student-written essays and LLM-generated text.
- To identify LLM artifacts that differentiate them from human writing.
- To contribute to addressing concerns about the potential impact of LLMs on education and plagiarism.

Approach:

- **Dataset:** A mix of student-written essays and essays generated by various LLMs.[Link]
- **Model Building:** Utilizing Natural Language Processing (NLP) techniques.
- **Feature Engineering:** Extracting linguistic and semantic features from the essays.
- **Machine Learning Algorithms:** Employing classification models to differentiate between student and LLM essays.



Deliverables

- **Trained Model:** A machine learning model capable of distinguishing between student and LLM-generated essays.
- **Feature Analysis:** Identification of specific linguistic or semantic features differentiating the two types of essays.
- **Documentation:** Report outlining the model's architecture, methodologies, and findings.

Evaluation Methodology:

- **Metric:** Accuracy, Precision, Recall, F1 Score for model performance.
- **Cross-Validation:** Testing the model on unseen data for generalizability.
- **Feature Importance:** Analyzing key features contributing to the model's classification.