

Mirza Akhi Khatun, PhD student at the University of Limerick,

**Project Title:**

**MCQ Question Answering with OpenAI ChatGPT-4 Model in the Kaggle LLM Science Exam**

**High-Level Project Description:**

This project aims to solve the Kaggle Large Language Models (LLMs) Science Exam problem by using OpenAI's language models, such as GPT-3.5 Turbo and GPT-4, to answer multiple-choice questions. The primary goal is to improve the model's proficiency in answering science exam questions.

LLMs, such as GPT-3.5 Turbo and GPT-4, are advanced Artificial Intelligence (AI) systems trained to recognize and respond to various prompts like humans. For example, the models are trained and evaluated on datasets containing science exam questions such as: "What is the powerhouse of the cell? A) Nucleus B) Ribosome C) Mitochondria D) Endoplasmic Reticulum." The correct answer is C) Mitochondria. Using the GPT-3.5 Turbo model, the project starts with a straightforward approach of asking the whole question directly to the model and having the model reply with the correct answer, achieving 61% accuracy.

Prompt engineering techniques, which involve carefully crafting the input text to guide the model in generating more accurate and relevant responses, improve accuracy by up to 73%. For example, instead of asking, "What is the powerhouse of the cell?" The prompt is modified to: "In the context of cellular biology, identify the powerhouse of the cell from the following options: A) Nucleus, B) Ribosome, C) Mitochondria, D) Endoplasmic Reticulum." This more detailed prompt helps the model focus on the specific context and options. As a result, the model answers questions more accurately.

As part of this project, we also focus on enhancing model accuracy and gaining valuable insights into LLMs. We utilize the dataset to access the OpenAI API using the GPT-4 model with Optuna optimization. This allows hyperparameters such as temperature to be fine-tuned, resulting in 86% accuracy. We demonstrate state-of-the-art performance for sophisticated language models by combining iterative optimization, prompt engineering, and advanced model capabilities.