# RAG EngSci Tool

Hanhee Lee

March 10, 2025

## 1 Introduction

Engineering Science at the University of Toronto is a one of the hardest undergraduate programs in Canada. Specfiically, the program is designed to learn a breadth of courses from all areas of engineering, including civil, software, biomedical, etc in a rigourous manner. As a result, I have seen first hand (including myself) many students stuggling with the program, especially those who did not have a strong background in math and physics or enroll in IB/AP programs in high school. To help these students, I am proposing a tool based on Retrieval Augmented Generation (RAG) that can help students learn the material in a more efficient manner. Everyone in this day in age is familiar with large language models (LLMs), such as the ones produced by OpenAI and Google. However, these models are general-purpose models, and are not tailored to a specific niche. This is why I am proposing a RAG model that is specifically tailored to the Engineering Science program, which will take in knowledge from past exams/midterms, lecture notes of recurring professors, and textbooks used in the program to generate a model that can help students learn the material.

## 2 Objectives

- **Curate Data:** Collect data from past exams/midterms using web scaping tools.
- **Train Model:** Train a RAG model using the data collected.
- **Evaluate Model:** Evaluate the model using a variety of metrics, including BLEU score, ROUGE score, etc.
- **Deploy Model:** Deploy the model as a web application that students can use to learn the material.
- **User Study:** Conduct a user study to see if the model is actually helping students learn the material.

## 3 Background

## 4 Methodology

### 4.1 Data

### 4.2 Evaluation

### 4.3 Model

RAGFlow is an open-source implmentation of a RAG model that will be used in this project.

## 5 References