Background of the Study

The rapid growth of digital learning technologies has transformed how students access and engage with information. In higher education, online platforms, e-books, and lecture recordings have become indispensable resources for both faculty and students, particularly following the COVID-19 pandemic (Lucero et al., 2022). Despite this shift, a recurring challenge remains: while students can access vast amounts of information, they often struggle to extract contextually accurate insights from lecture notes and supplementary materials (Rosales, 2021).

Artificial intelligence (AI) has been integrated into education via chatbots and virtual assistants that can support student learning, enhance engagement, and act as study companions; however, they often face limitations in contextual understanding, reliability, and accuracy (Labadze, Grigolia, & Machaidze, 2023). A key problem is that large language models (LLMs) are prone to **hallucinations**, producing content that appears factually correct but is not grounded in reliable sources (Rawte et al., 2023).

To address hallucinations, **Retrieval-Augmented Generation (RAG)** was introduced by Lewis et al. (2020), combining document retrieval with generative AI to ensure outputs are based on factual knowledge (Lewis et al., 2020). While Retrieval-Augmented Generation (RAG) enhances factual grounding, it continues to face limitations in **document chunking**, **retrieval effectiveness**, and **prompt optimization**. Studies have shown that semantic chunking often increases computational costs without consistently improving retrieval quality (Qu, Tu, & Bao, 2024), that optimal chunk size depends heavily on dataset and task requirements (Bhat, Rudat, Spiekermann, & Flores-Herr, 2025), and that poorly structured prompts can hinder accuracy compared to modular and context-aware designs (Trisciuoglio, 2025). These challenges are particularly evident when applied to academic lecture notes, where nuanced context and semantic coherence are essential.This thesis proposes an **improved RAG framework** specifically designed for educational applications, with three main innovations:

1. **Improved Chunking** – refining how documents are divided into semantically meaningful sections to avoid context loss and improve answer accuracy
2. **Enhanced Retrieval Techniques** – integrating advanced retrieval models such as ColBERT to ensure more precise and contextually relevant document selection
3. **Optimized Prompt and Retrieval Integration** – designing structured prompts that guide the model to use retrieved knowledge effectively, minimizing hallucinations and improving response reliability

The beneficiaries of this study are primarily **students in higher education**, who will gain access to AI-driven tools that provide accurate, grounded, and context-aware assistance when studying lecture notes. Faculty members may also benefit through improved student preparedness and reduced time spent clarifying repetitive queries. At a broader scale, this study contributes to the field of educational technology by demonstrating how **targeted improvements in RAG can enhance AI-powered academic support systems**.

By addressing the limitations of existing RAG systems through improved chunking, retrieval, and optimized prompt integration, this research fills a critical gap in the intersection of **AI, education, and natural language processing**.

https://www.researchgate.net/publication/358604030_Assessment_of_E-Learning_Readiness_of_Faculty_Members_and_Students_in_the_Government_and_Private_Higher_Education_Institutions_in_the_Philippines

https://www.researchgate.net/publication/380923616_An_Empirical_Study_on_Knowledge_Access_and_Utilization_of_Electronic_Resources_by_University_Students_in_Philippines

https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-023-00426-1

https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

https://www.researchgate.net/publication/386168512_Is_Semantic_Chunking_Worth_the_Computational_Cost

https://www.researchgate.net/publication/392167724_Rethinking_Chunk_Size_For_Long-Document_Retrieval_A_Multi-Dataset_Analysis

https://medium.com/@giuseppetrisciuoglio/prompt-engineering-for-rag-advanced-strategies-to-maximize-llm-accuracy-part-1-3e283230f2c0