Capstone Project Proposal

1. Proposed Title

“Retrieval-Augmented Generation (RAG) for Enterprise-Grade Chatbots with Hallucination Control and Reinforcement Learning-Based Retrieval Optimization”

---

2. Methodology

The project will follow these major steps:

(a) Knowledge Base (KB) Setup

- Collect domain-specific documents (e.g., HR policies, university regulations, or legal sections).
- Clean and preprocess text, split into chunks, and embed using a transformer-based embedding model.
- Store embeddings in a vector database (FAISS/Chroma).

(b) RAG Pipeline Construction

- Implement a retrieval-augmented generation chatbot.
- For each query, retrieve top-k relevant chunks from the KB.
- Pass retrieved chunks + query to an LLM for grounded response generation.

(c) Reinforcement Learning for Retrieval Optimization *(Novelty #1)*

- Train a retrieval policy to dynamically select the optimal number of documents (k) to balance **accuracy** and **cost (tokens)**.
- Use reward signals based on response correctness and retrieval efficiency.

(d) Hallucination Detection Module *(Novelty #2)*

- Compare generated responses with retrieved evidence to detect unsupported or fabricated content.
- Techniques: embedding similarity, natural language inference (NLI), or SelfCheckGPT-style re-generation.

(e) Evaluation & Testing

- Measure performance using both automated metrics (BLEU, ROUGE, F1, retrieval accuracy) and human evaluation.
- Compare baseline RAG chatbot vs. RL-optimized + hallucination-controlled version.

---

3. Expected Outcome

- A working **enterprise-grade chatbot prototype** (Streamlit/Gradio app) capable of answering domain-specific queries accurately.
- **Reduced hallucination rate**, demonstrated via a hallucination detection module.
- **Improved cost-efficiency** and accuracy through reinforcement learning optimization of retrieval.
- A **capstone report** highlighting:
    - Research foundation (2024 RAG & FACTS papers).
    - Novel contributions (RL retrieval + hallucination detection).
    - Industry applicability (secure, scalable, enterprise-ready chatbot).