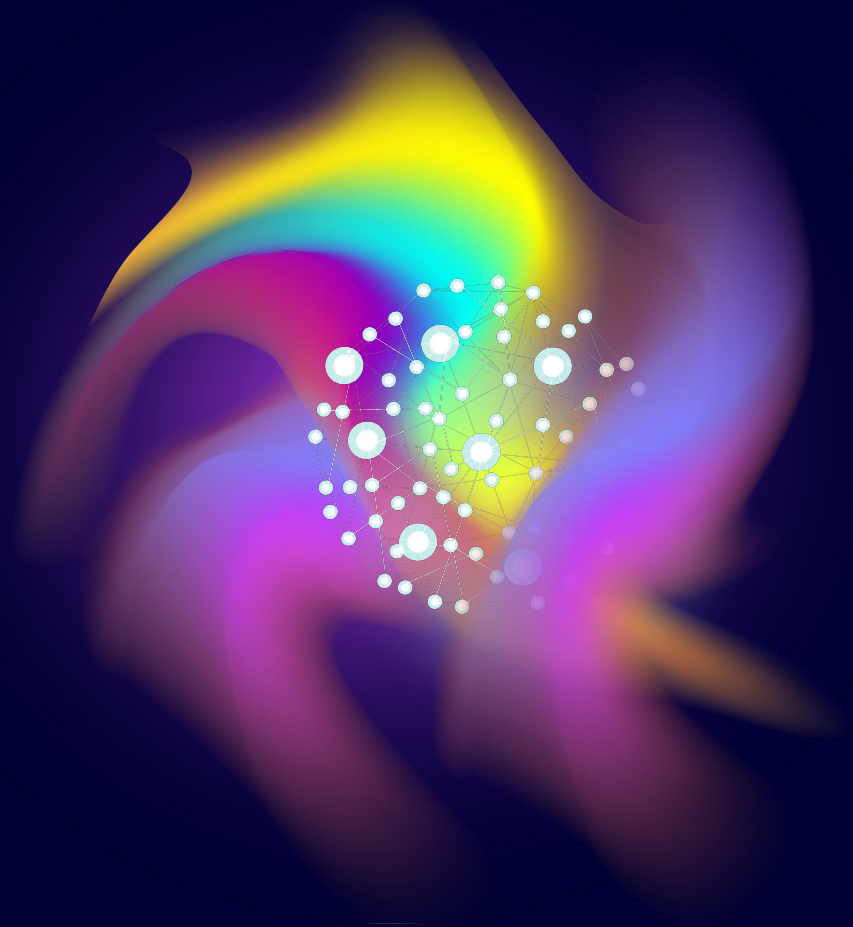
SPU ChatBot

Administration ChatBot for new students



Owais Hilal 4200051

Dr. Mouhib Almoukkari Eng.Aya Alaswad

Table of contents

01

Introduction

The rapid growth of digital information in universities has made it increasingly difficult for students to quickly locate the exact details they need. Academic regulations, course descriptions, exam rules, and general university guidelines are often spread across multiple PDFs, webpages, and documents. Students frequently experience frustration navigating these resources, especially when they need fast answers during registration periods, exam preparation, or course selection. An AI-powered chatbot provides an efficient solution by delivering instant, accurate responses based strictly on trusted university documents. This system reduces confusion, improves student experience, and ensures reliable access to important information.

02

Problem Statement

Universities store large amounts of critical academic information, but students often struggle to find what they need due to scattered and unstructured document formats. Searching through long PDFs or outdated webpages consumes valuable time, and students may rely on incorrect answers from peers or social media. Traditional search functions cannot understand context, so users must manually sift through many pages to find the relevant information. There is currently no centralized and intelligent system capable of understanding natural language questions and retrieving the most accurate and relevant answers directly from university documents.

Key Problems

Academic information is scattered across many PDFs, documents, and webpages.

Students must manually search multiple sources, wasting time.

Traditional keyword search cannot understand meaning or context.

Information may be outdated or difficult to locate quickly.

Students often rely on unreliable sources such as peers or social media.

No single platform provides fast, accurate, AI-driven answers. innovative technologies

03

Objectives

The primary objective of this project is to develop a reliable and intelligent AI-powered university information chatbot capable of understanding and answering student queries accurately and efficiently. Specifically, the project aims to:

1. **Accurate Retrieval:** Enable the chatbot to provide responses grounded in verified university documents.
2. **Natural Language Understanding:** Allow the system to comprehend a wide variety of question formats posed in natural language.
3. **Scalability:** Design the system to easily incorporate new data annually or as university policies change.
4. **User-Friendly Interface:** Provide an intuitive chat interface for students and a simple admin interface for uploading or updating documents.
5. **Efficiency:** Reduce the time students spend searching for information and minimize repetitive inquiries directed at staff.
6. **Reliability:** Ensure that answers are consistent, factual, and up-to-date.
7. **Integration:** Seamlessly combine vector databases, embeddings, and LLMs for a robust RAG-based solution.

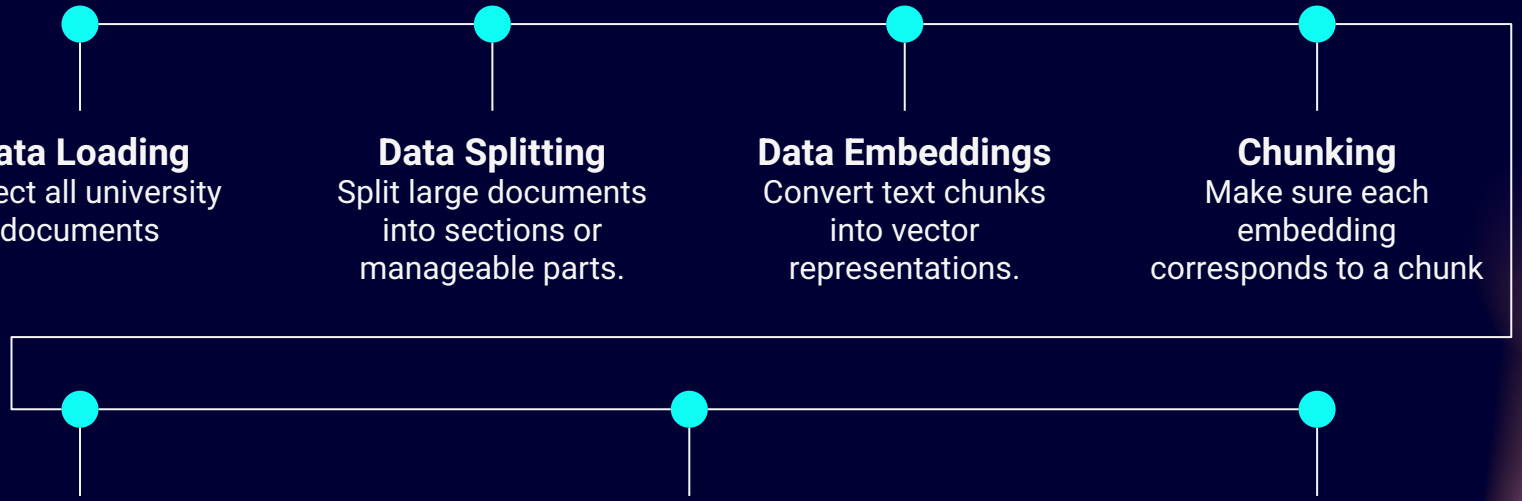
This set of objectives ensures the project not only functions effectively but also adds significant value to the student experience and university administrative efficiency.

04

Methodology

This project leverages LangChain to efficiently manage the RAG (Retrieval-Augmented Generation) workflow, integrating document processing, embeddings, vector stores, retrieval, and LLM generation. LangChain provides a modular and maintainable framework that orchestrates each step of the chatbot pipeline.

Project timeline



05

System Architecture

The chatbot system is designed as a microservices architecture, which allows each step of the pipeline to run independently and communicate via APIs. This ensures scalability, maintainability, and flexibility for future updates.

•Data Ingestion Service
Handles uploading and updating of university documents. Supports multiple formats including PDFs and text files.

• Preprocessing Service
Extracts text, cleans, and chunks documents. Runs independently to allow parallel processing of multiple files.

•Embedding Service
Converts chunks into vector embeddings using Sentence-Transformers.

•Vector Store Service
Maintains the Milvus or Qdrant database, handles storage, retrieval, and indexing.

•Retrieval Service
Fetches relevant chunks from the vector store when a user submits a query.

•Question Answering Service
Integrates the retrieved chunks with the LLM to generate accurate and grounded answers.

Gantt Chart

Week	1	2	3	4	5	6	7	8	9	10
Data Loading										
Data Splitting										
Data Embeddings										
Chunking										
Vectorstore										
Retrieving										
Question Answering										
Frontend Development										
Testing & Documentation										

06

Conclusion

This project addresses the challenges students face when trying to access university information efficiently. By leveraging a modern RAG-based pipeline orchestrated through LangChain, the chatbot can provide accurate and reliable answers grounded in verified university documents. The microservices architecture ensures modularity, scalability, and maintainability, allowing the system to easily incorporate new data each academic year.

Key benefits include:

1. Rapid access to university information without searching multiple sources.
2. Reduction in repetitive queries handled by staff.
3. Scalable and maintainable AI-driven system.
4. Integration of vector databases, embeddings, and LLMs for accurate responses.
5. User-friendly interface for both students and administrators.

Overall, this project demonstrates a practical application of AI to enhance student experience and improve information accessibility within the university.

Thank You!!