

Chatbot for DBU

Lohitha Mahesh





Agenda



Introduction

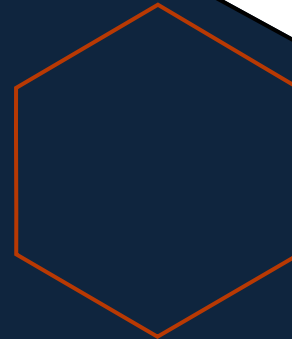
This project focuses on developing a chatbot specifically designed for the DBU community. The chatbot has been trained on the university's official catalog, enabling it to answer queries directly related to:

- University Schedule
- Admissions
- Academic Programs
- Financial Aid
- General Policies



Problem Statement

At DBU, students, faculty, and staff often struggle to get quick answers and support because information is scattered across multiple sources, making it harder to resolve queries efficiently and smoothly.



Data Collection

- DBU Calendar for SP25 Term: Provides details on important dates and deadlines for the term, such as registration, exams, and holidays.
- DBU Catalog: Contains comprehensive information on courses, programs, policies, and other academic details.

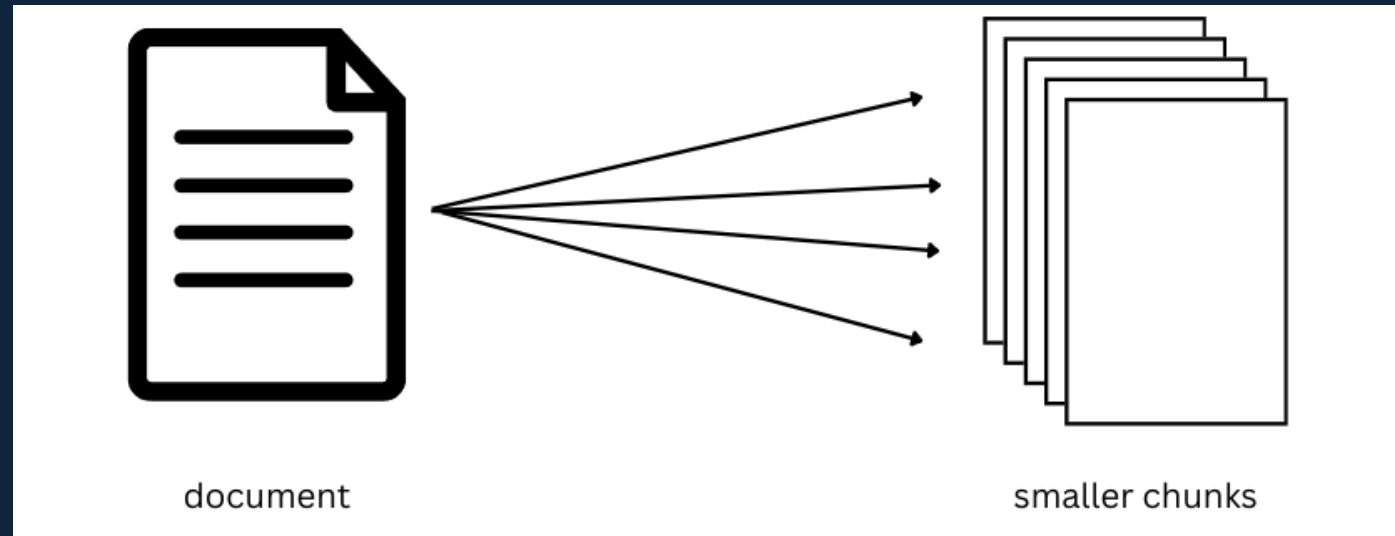


Preprocessing

- It extracts raw text from PDFs, converting unstructured data into a machine-readable format.
- Next, large documents are split into smaller sections for efficient processing and embedding.



Extract Text



Model Development

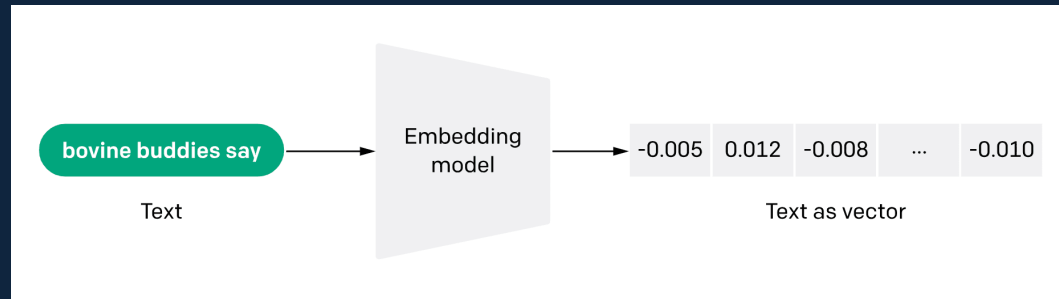
Embedding Creation

Example: Input Text (Chunk from pdfs):

Registration for Spring 2025 begins on January 10th and closes on January 20th. Classes start on January 25th.

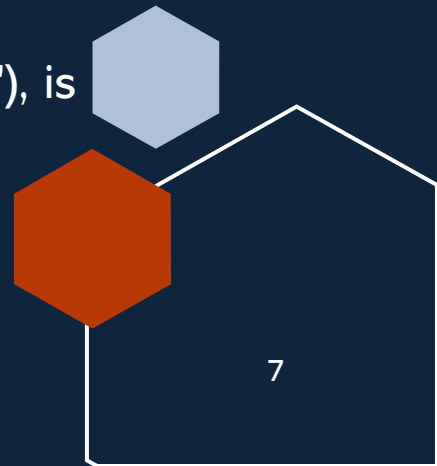
Steps:

1. Convert Text to Embeddings: The text is transformed into a vector representation using OpenAIEmbeddings.

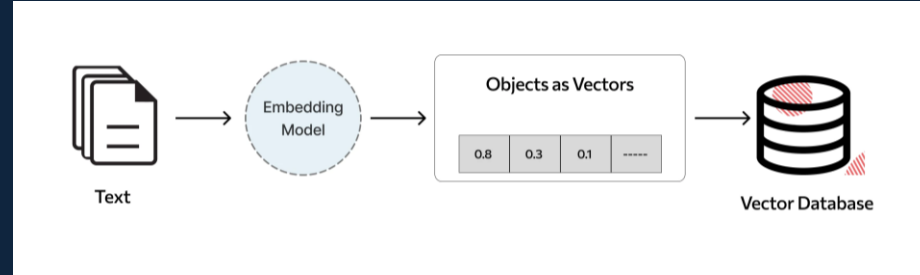


2. Store in FAISS Vector Database: This vector, along with its metadata (e.g., "Source: DBU Calendar"), is added to the FAISS database.

- Metadata example:
`{"source": "DBU Calendar", "chunk": "Registration Dates"}`

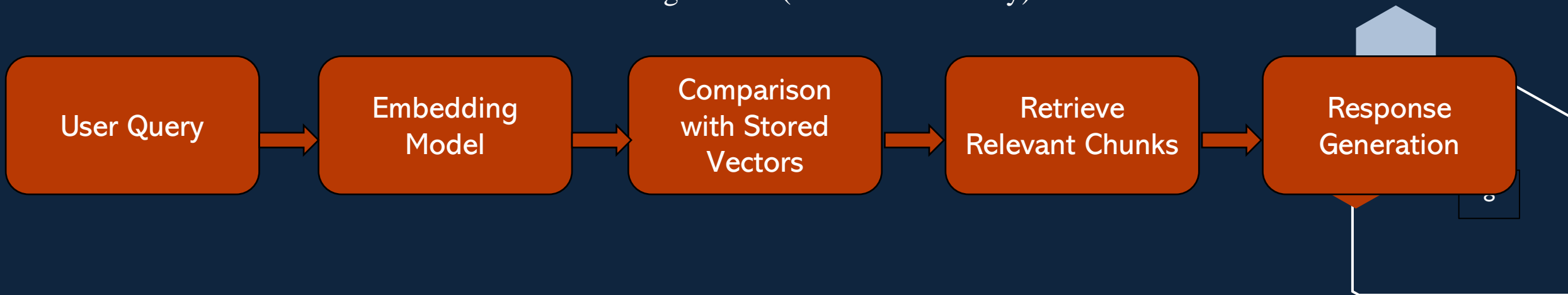


Model Development Cont.

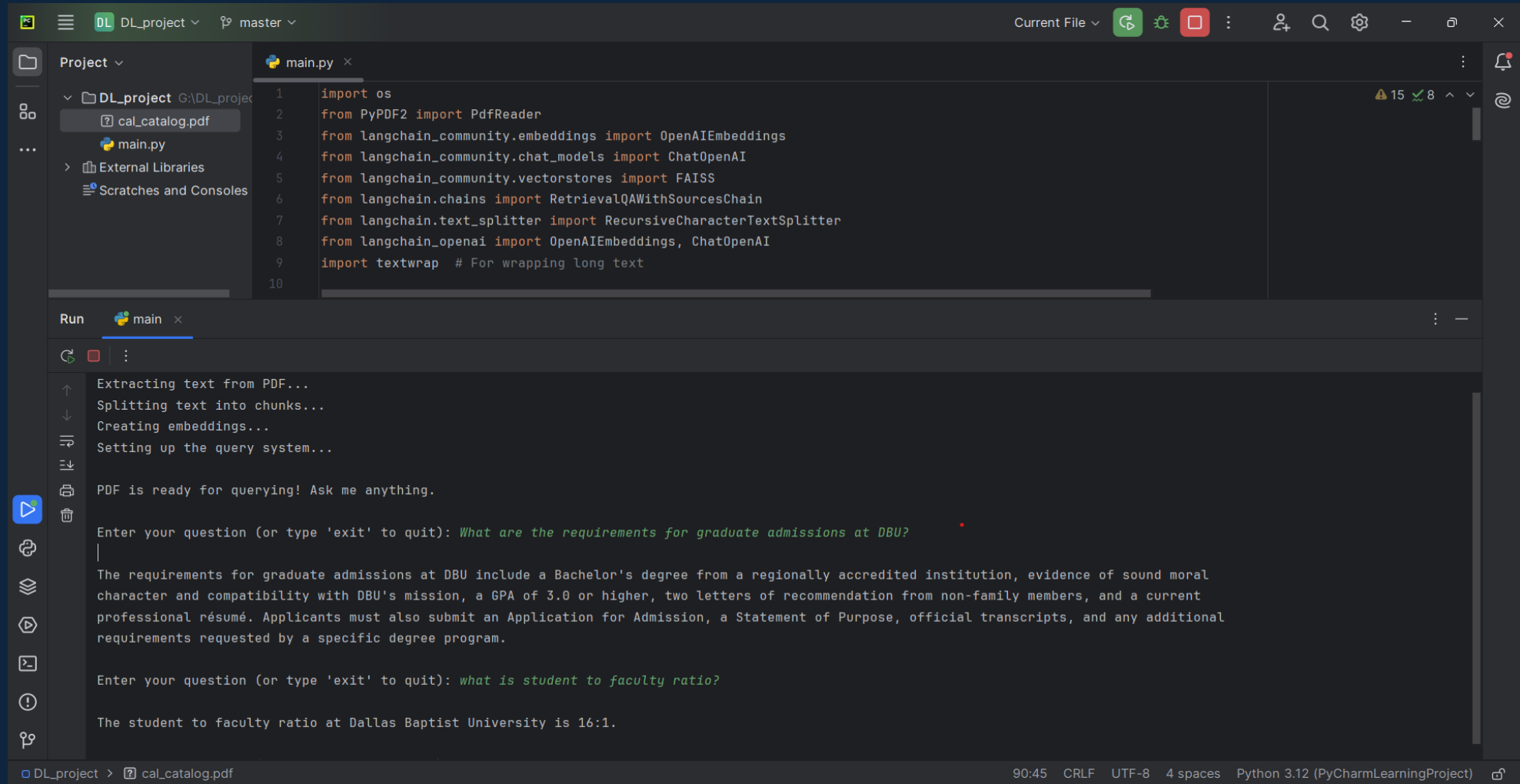


3. Retrieve Using Similarity Search:

- When a user asks: *"When does registration start for Spring 2025?"*, the chatbot converts the question into a query vector.
- The FAISS database retrieves the closest matching vectors (based on similarity).



Results



The screenshot displays the PyCharm IDE interface. The top toolbar includes icons for file operations, search, and settings. The left sidebar shows the project structure with 'DL_project' and 'cal_catalog.pdf'. The main editor window shows the code for 'main.py'.

```
1 import os
2 from PyPDF2 import PdfReader
3 from langchain_community.embeddings import OpenAIEmbeddings
4 from langchain_community.chat_models import ChatOpenAI
5 from langchain_community.vectorstores import FAISS
6 from langchain.chains import RetrievalQAWithSourcesChain
7 from langchain.text_splitter import RecursiveCharacterTextSplitter
8 from langchain_openai import OpenAIEmbeddings, ChatOpenAI
9 import textwrap # For wrapping long text
10
```

The bottom panel shows the 'Run' output for 'main.py'. The execution steps are: Extracting text from PDF..., Splitting text into chunks..., Creating embeddings..., and Setting up the query system... The output then displays the PDF content and the results of two queries.

PDF is ready for querying! Ask me anything.

Enter your question (or type 'exit' to quit): *What are the requirements for graduate admissions at DBU?*

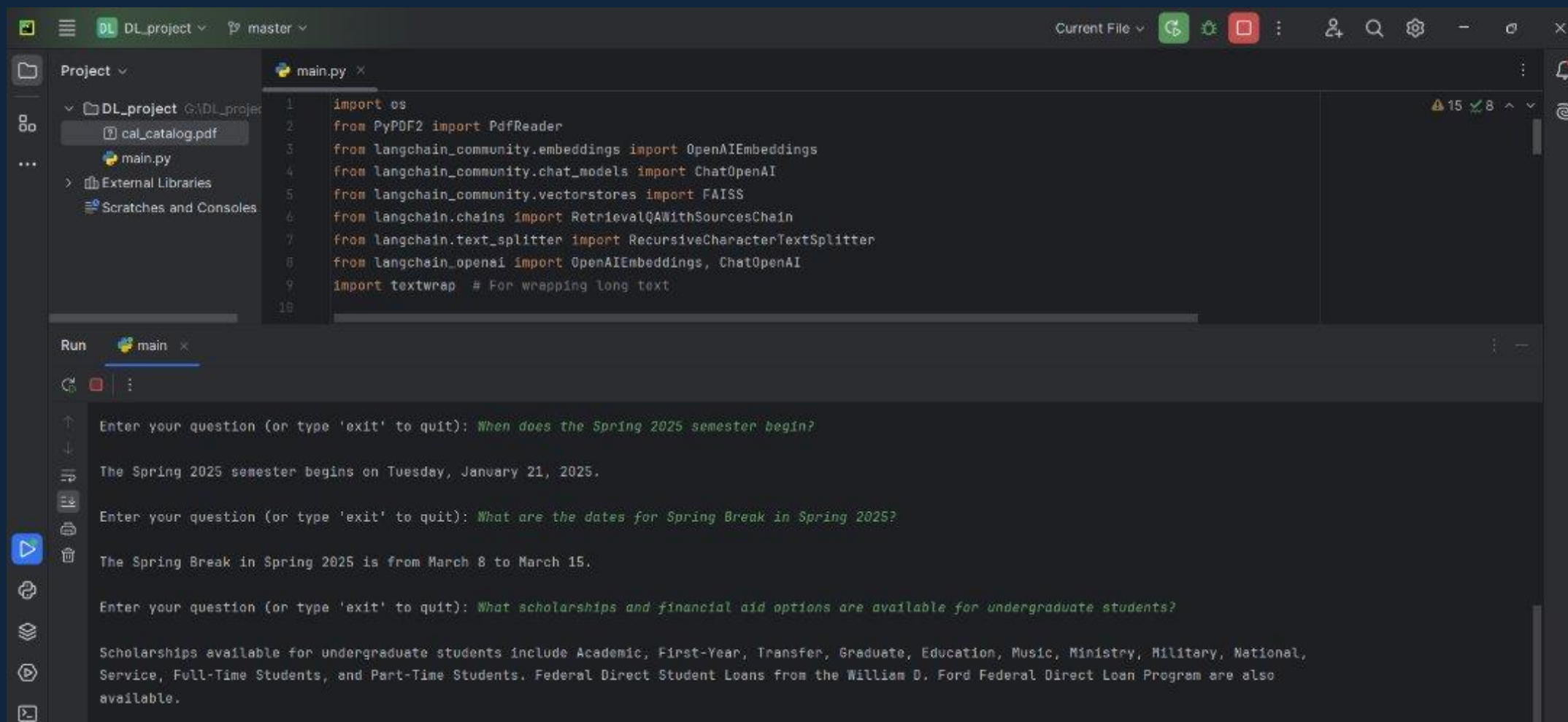
The requirements for graduate admissions at DBU include a Bachelor's degree from a regionally accredited institution, evidence of sound moral character and compatibility with DBU's mission, a GPA of 3.0 or higher, two letters of recommendation from non-family members, and a current professional résumé. Applicants must also submit an Application for Admission, a Statement of Purpose, official transcripts, and any additional requirements requested by a specific degree program.

Enter your question (or type 'exit' to quit): *what is student to faculty ratio?*

The student to faculty ratio at Dallas Baptist University is 16:1.

The status bar at the bottom indicates the file is 'cal_catalog.pdf' in the 'DL_project' directory, with settings for 90:45, CRLF, UTF-8, 4 spaces, Python 3.12, and the project name 'PyCharmLearningProject'.

Results Cont.



The screenshot shows a code editor with a file named `main.py` open. The code imports various libraries including `os`, `PyPDF2`, `langchain_community`, `langchain`, `langchain_openai`, and `textwrap`. The code is designed to process a PDF file (`ca_catalog.pdf`) and answer questions based on its content.

```
1 import os
2 from PyPDF2 import PdfReader
3 from langchain_community.embeddings import OpenAIEmbeddings
4 from langchain_community.chat_models import ChatOpenAI
5 from langchain_community.vectorstores import FAISS
6 from langchain.chains import RetrievalQAWithSourcesChain
7 from langchain.text_splitter import RecursiveCharacterTextSplitter
8 from langchain_openai import OpenAIEmbeddings, ChatOpenAI
9 import textwrap # For wrapping long text
10
```

The execution results are displayed in the Run console, showing a series of questions and answers:

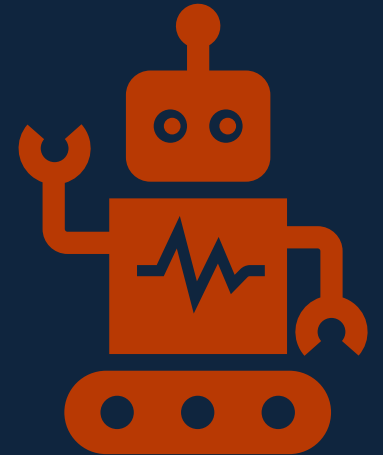
```
Run main x
Enter your question (or type 'exit' to quit): When does the Spring 2025 semester begin?
The Spring 2025 semester begins on Tuesday, January 21, 2025.
Enter your question (or type 'exit' to quit): What are the dates for Spring Break in Spring 2025?
The Spring Break in Spring 2025 is from March 8 to March 15.
Enter your question (or type 'exit' to quit): What scholarships and financial aid options are available for undergraduate students?
Scholarships available for undergraduate students include Academic, First-Year, Transfer, Graduate, Education, Music, Ministry, Military, National, Service, Full-Time Students, and Part-Time Students. Federal Direct Student Loans from the William D. Ford Federal Direct Loan Program are also available.
```

Summary

- The chatbot is developed to enhance user experience. It leverages AI to deliver accurate and real-time responses.

The solution addresses key challenges and provides a scalable foundation for future advancements.

- Chatbots are widely adopted in sectors like e-commerce, healthcare, and banking, offering tailored solutions for customer support, lead generation, and internal operations.



Thank you :)

