Chatbot for DBU

Lohitha Mahesh

Agenda

- Introduction
- Problem Statement
- Data Collection & Preprocessing
- Model Development
- Result & Insights

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

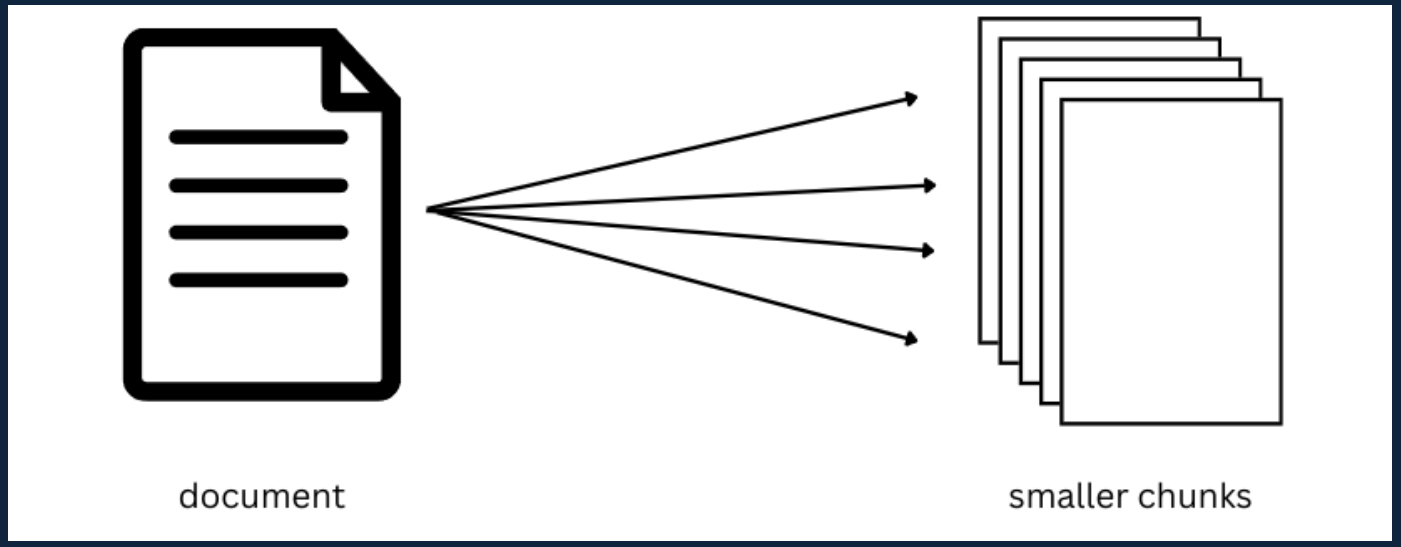
At DBU, students, faculty, and staff often struggle to get quick answers and support because information is scattered across multiple sources, making it harder to resolve queries efficiently and smoothly.

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.

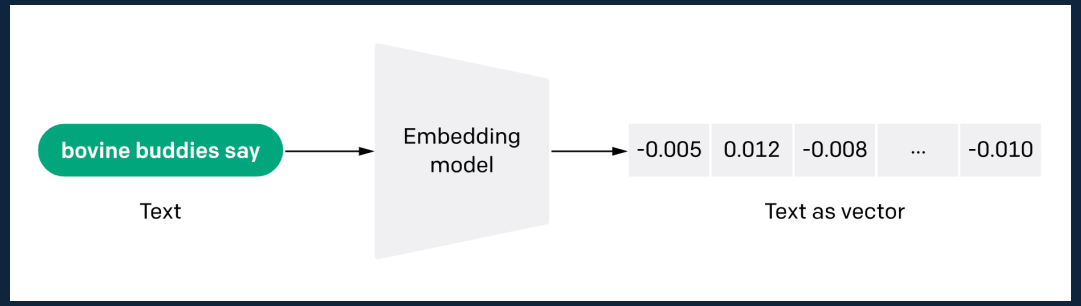


Model Development

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.

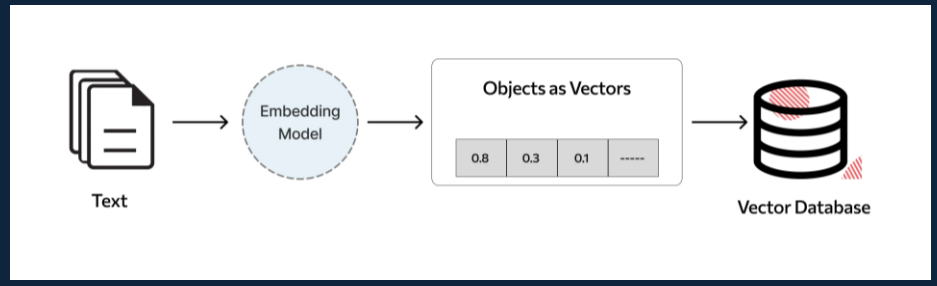
Steps:
1. Convert Text to Embeddings: The text is transformed into a vector representation using OpenAIEmbeddings.



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

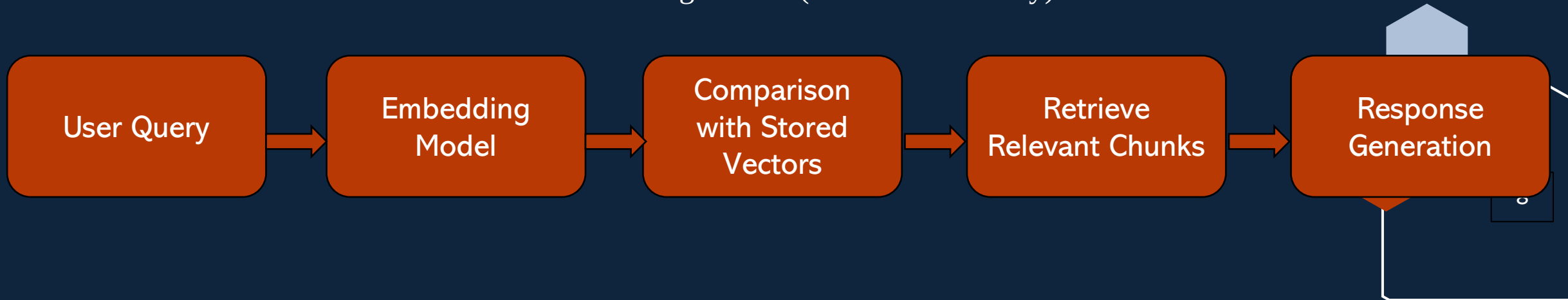
- Metadata example:
  `{"source": "DBU Calendar", "chunk": "Registration Dates"}`

Model Development Cont.

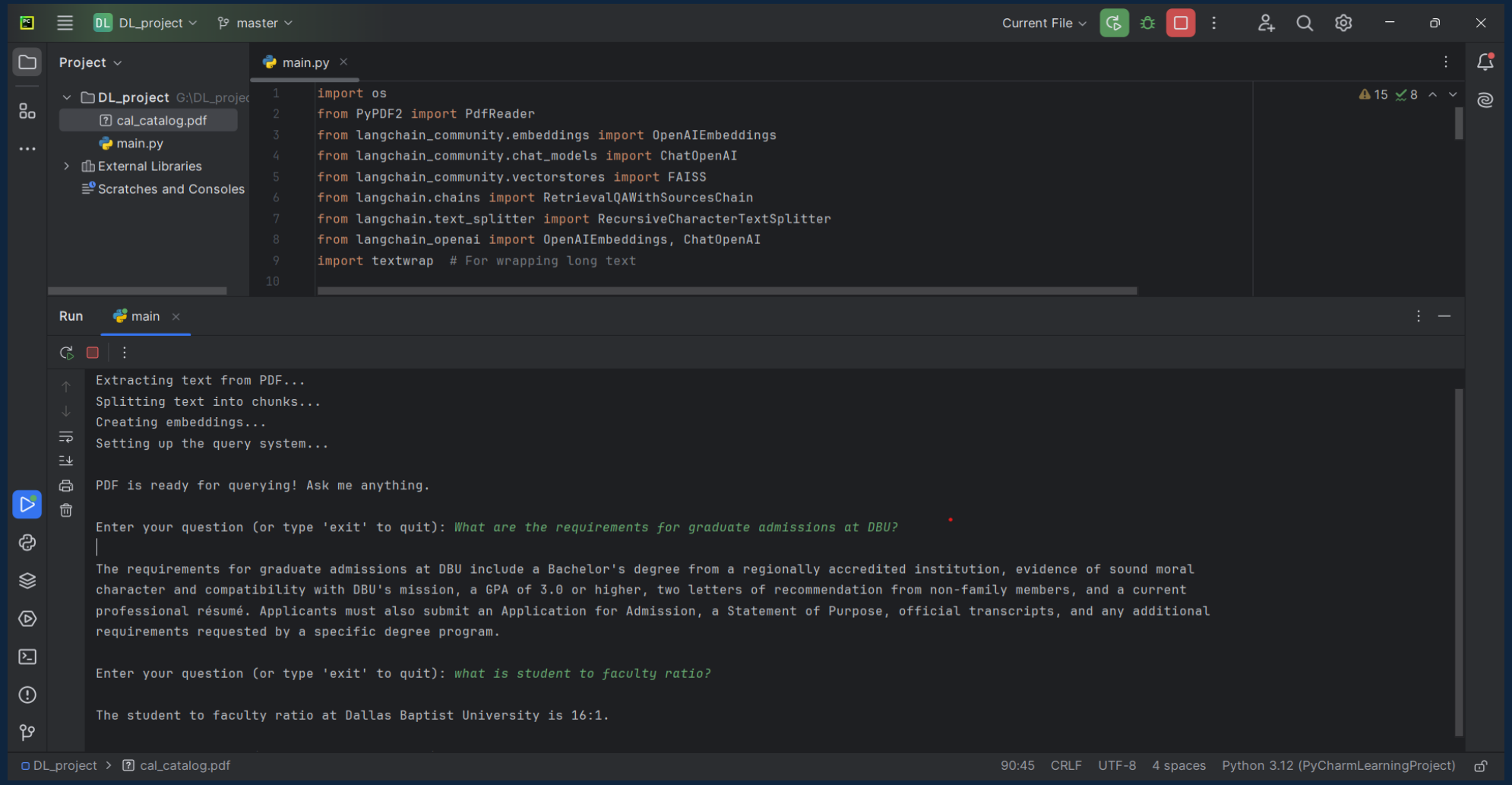


3. Retrieve Using Similarity Search:

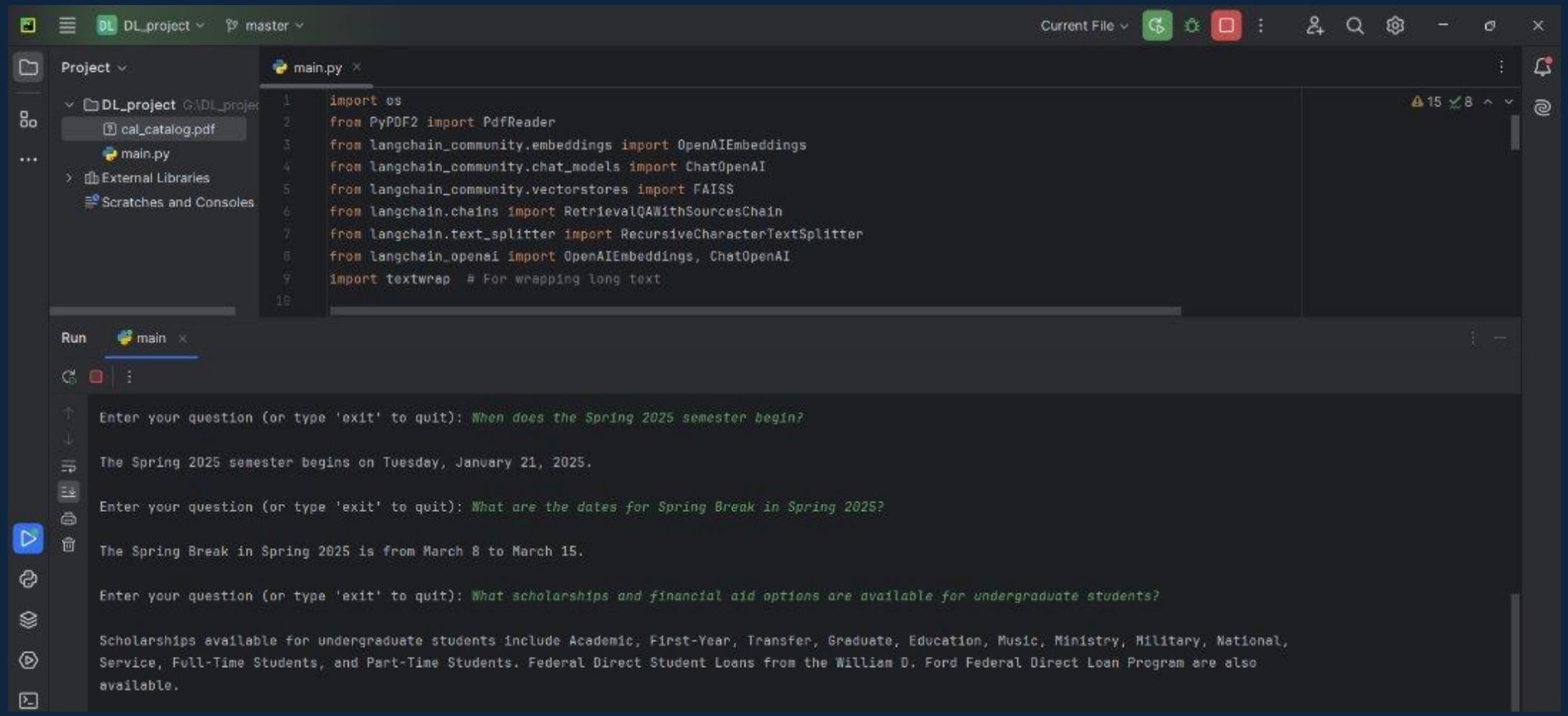
- When a user asks: *"When does registration start for Spring 2025?"*, the chatbot converts the question into a query vector.
- The FAISS database retrieves the closest matching vectors (based on similarity).



Results


DL_project
master
Current File
Project
main.py
DL_project G:\DL_projec
cal_catalog.pdf
main.py
External Libraries
Scratches and Consoles
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
15 8
Run
main
Extracting text from PDF...
Splitting text into chunks...
Creating embeddings...
Setting up the query system...
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.
DL_project > cal_catalog.pdf
90:45 CRLF UTF-8 4 spaces Python 3.12 (PyCharmLearningProject)

Results Cont.



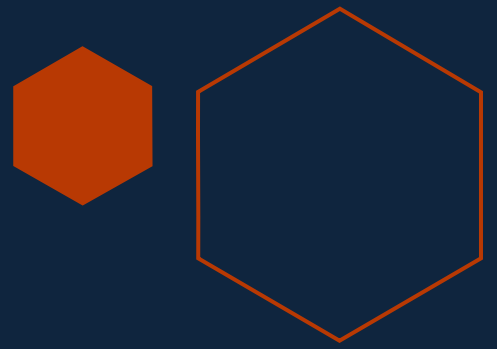
Summary

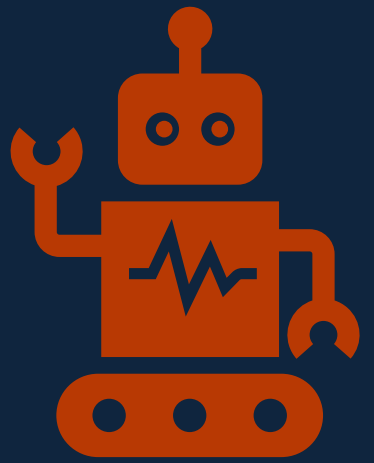
- The chatbot is developed to enhance user experience. It leverages AI to deliver accurate and real-time responses.

  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 :)