

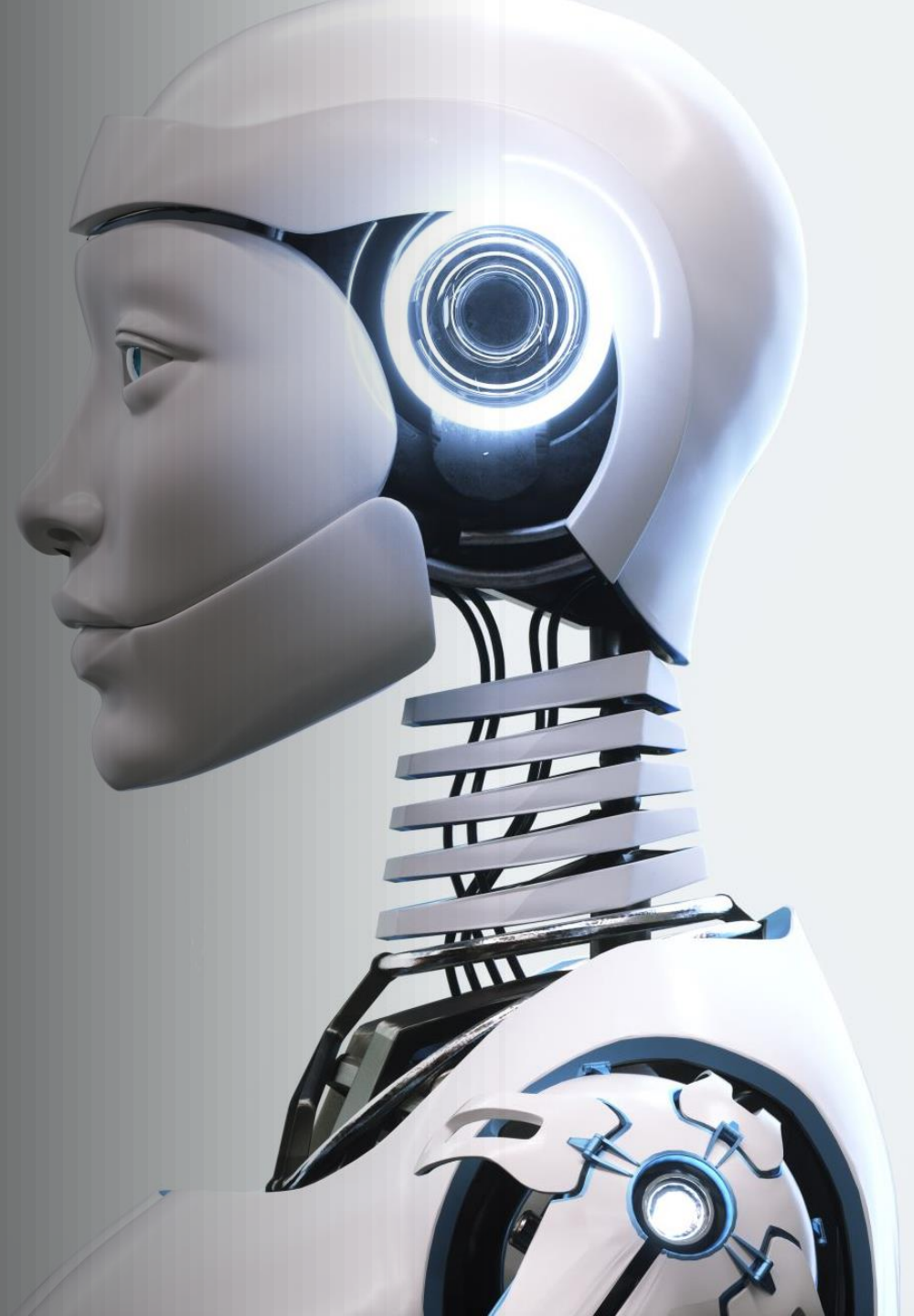
# Context-Based ChatBot using RAG System

---

Submitted by:

Keerthana Goka

Kunal Malhan



The background image shows two white, humanoid robots sitting opposite each other at a small, light-colored table. They are in a room with a warm, orange-brown wall. The robot on the left is seated on a green cushioned chair, and the robot on the right is on a similar green chair. On the table between them are several small, white, cup-like objects and a small vase with a plant. The robots have a sleek, modern design with visible joints and a human-like posture. The entire scene is framed by a white, hand-drawn style border.

# Project Overview

---

Thoughts behind chatbot

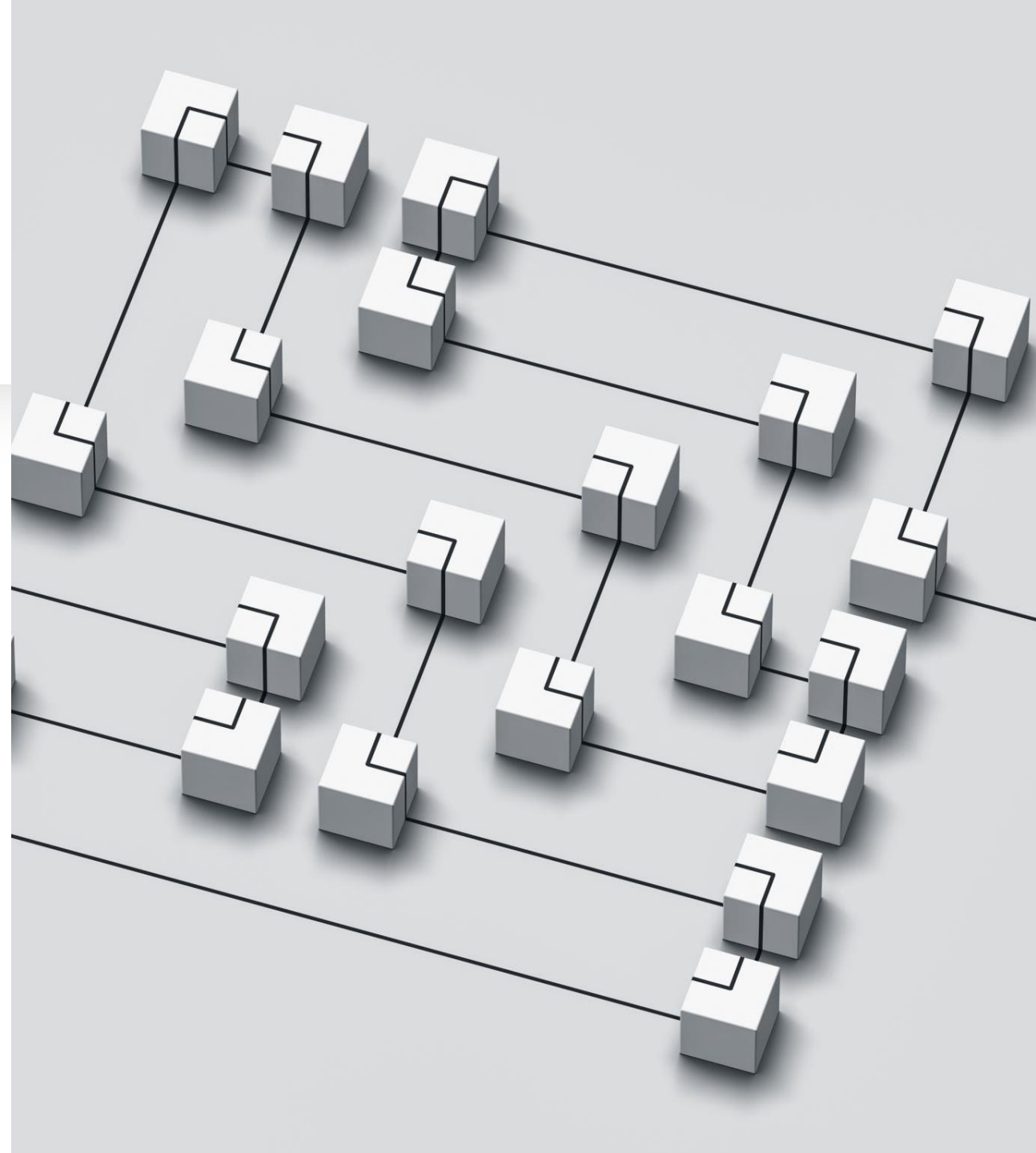
# Problem Description

- This project will aim to use Retrieval Augmented Generation (RAG) system to get domain specific knowledge with base of Large Language Model (LLM) like llama2 or similar.
- The challenge for implantation will reside in understanding the user question and the responding appropriate information and developing ethical guidelines for interaction
- All communication and input/outputs will be limited to English language.



# Scope of the Project

- For RAG system implementation:
  - Utilizing LLM models through Lang-Chain or Hugging-Face library
  - Loading contextual data files in form of text-chunks and store as embeddings in vector database.
  - Based on context of question retrieval of relevant data chunks.
  - Generate response for user.
  - Save response as chat-history.
- Implementation of user interface for chatbot.







# Skills and Tools Required

- Python
- pandas, numpy
- LangChain library
- Hugging Face
- Streamlit



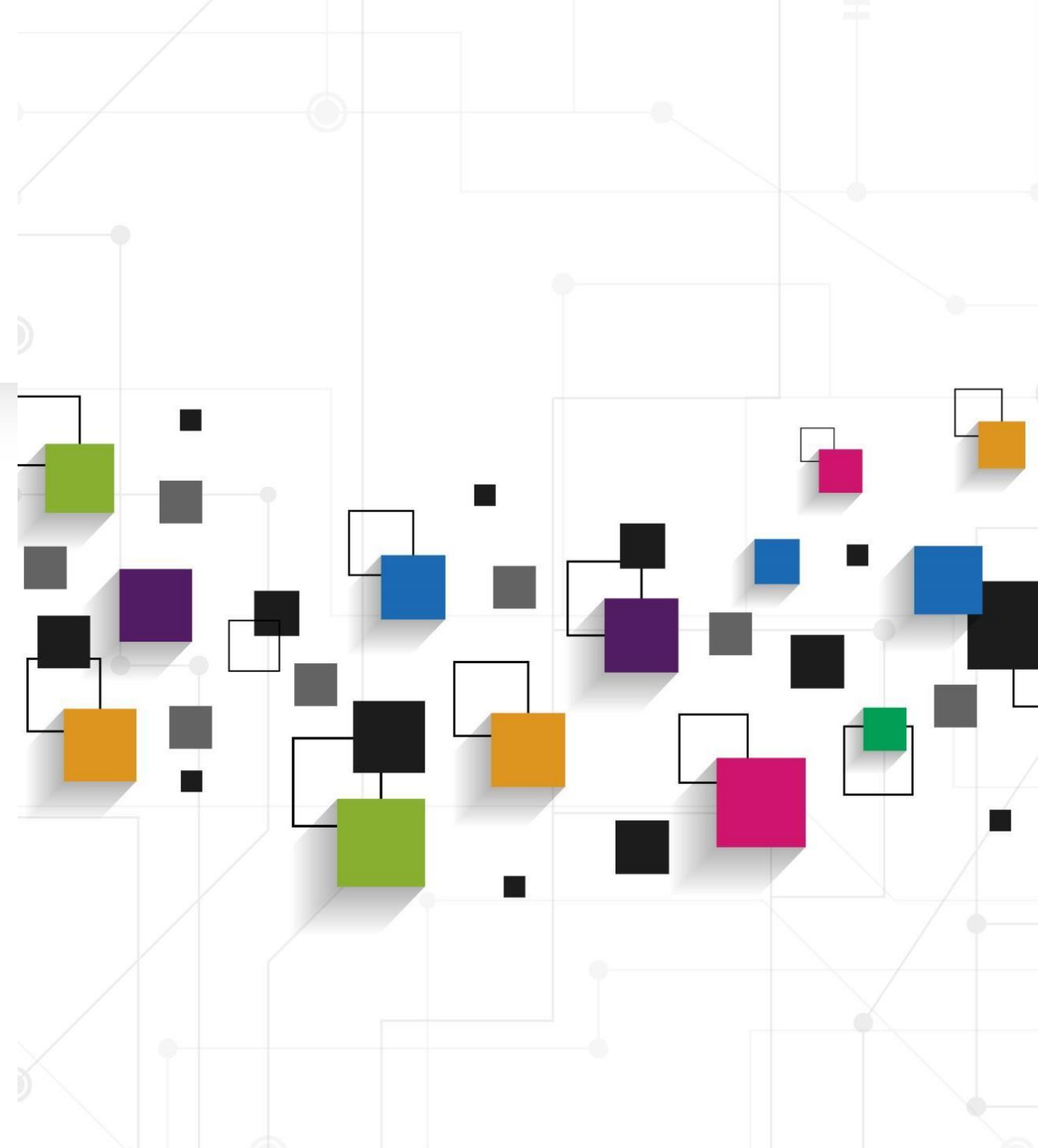
# Kanji Chatbot

---

Insights at Your Fingertips – Just Upload and Ask.

# Data Processing Pipeline

- For model training on contextual dataset:
  - Upload files
    - Virtually any text content in PDF file can be source of information for the system.
  - Create documents from files
  - Split documents into chunks
  - Create embeddings for chunks
  - Storing chunks and embeddings in vector database (ChromaDB)
- For conversational Q&A:
  - Create embeddings for user message
  - Search for relevant chunks/documents from vector database
  - Generate response and display to user
  - Add response to context for memory generation.



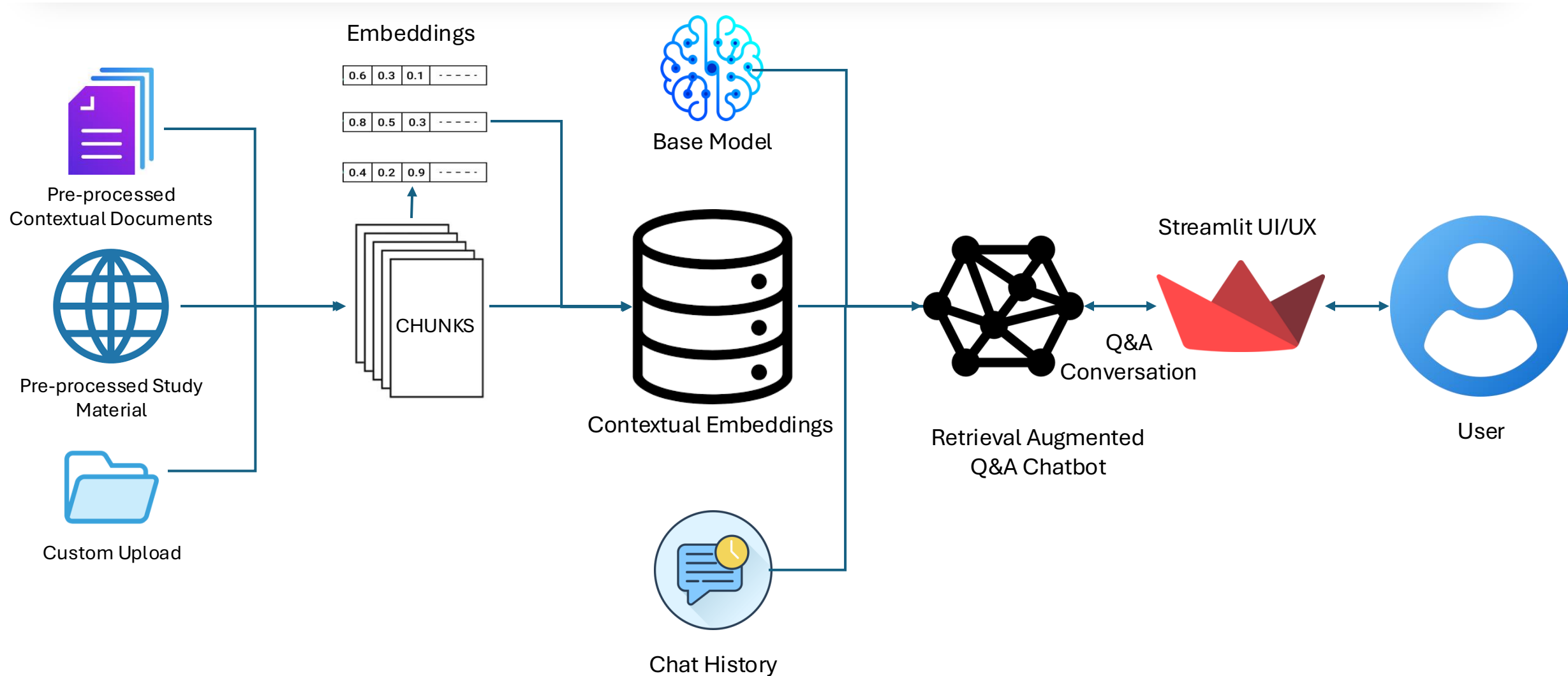


# Methodology

- Our methodology includes the use of natural language processing, large language models and then adding contextual layer above the base model. Step by step methodology is as follows:
  - Download pre-trained model through Ollama
    - We will be using “Gemma2-9b” as base model. Advantage of using this includes its open source, easy integration, and small (need relatively less compute).
  - Setting up base model using langchain
    - langchain-community is open-source library that will be used to setup the complete layers with base model.
  - Creating Retrieval Augmented Generation (RAG) layer
  - Creating web-based user interface for chatbot using Streamlit framework



# Kanji Chatbot Architecture



# UI/UX Features of Kanji Chatbot



Chatbot access through  
Login or Anonymously



User based chat  
memorization in case of  
login



Multiple  
departmental/use-cases  
fixed/preloaded chat  
windows



Flexibility to load your own  
pdf documents



# Implementation and demo

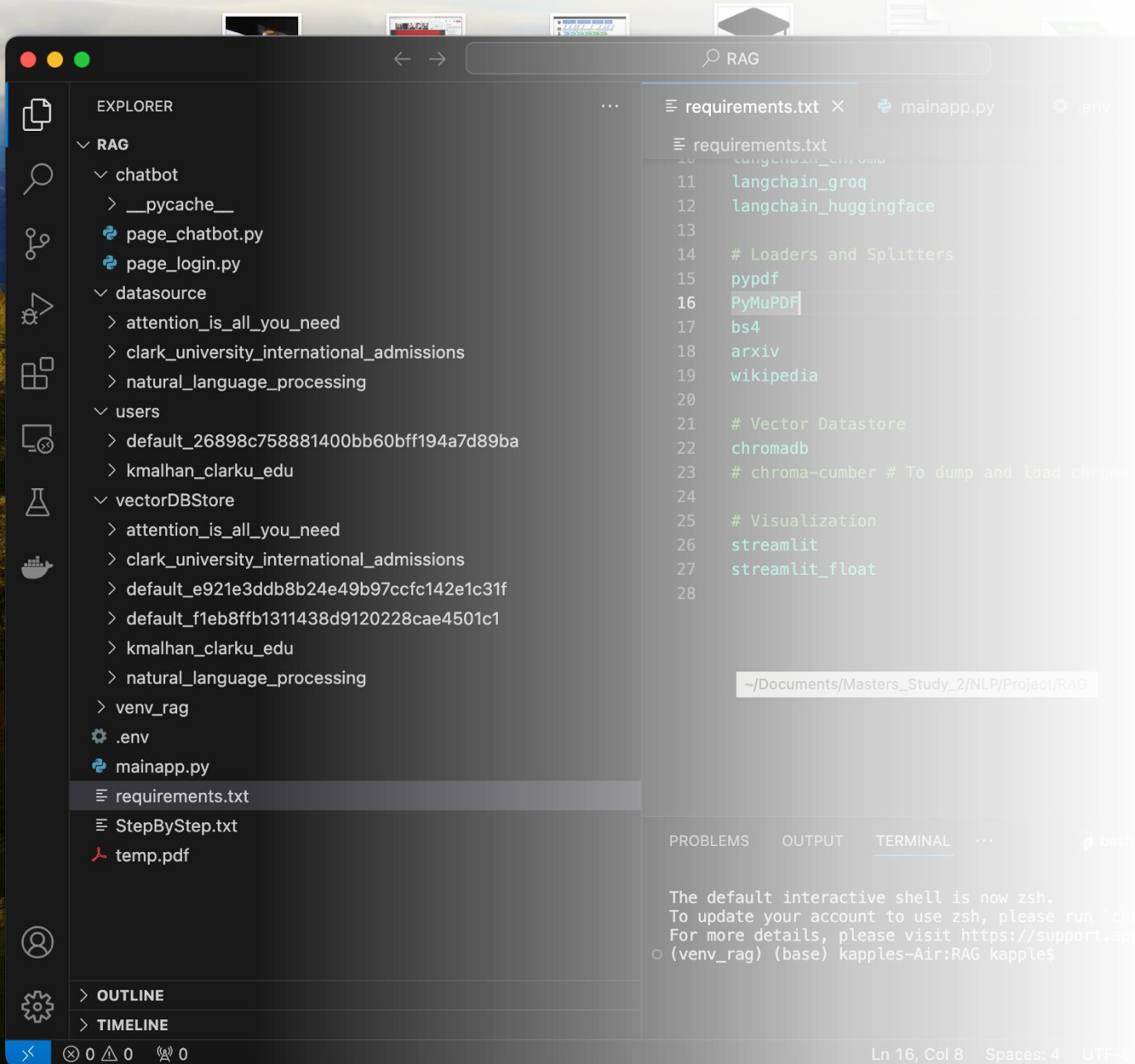
---

Kanji Chatbot

# Libraries used

- Python Environment
  - ipykernel
  - python-dotenv
- Langchain
  - langchain
  - langchain-openai
  - langchain-community
  - langchain-text-splitters
  - langchain\_chroma
  - langchain\_groq
  - langchain\_huggingface
- Loaders and Splitters
  - pypdf
  - PyMuPDF
  - bs4
  - arxiv
  - wikipedia
- Vector Datastore
  - chromadb
- Visualization
  - streamlit
  - streamlit\_float





# Platform

- Visual Studio
- Important files

# Screenshots

## Kanji Chatbot

User ID (Clark Email address):

Password (Your firstname):



Login

Continue without login

## Kanji Chatbot

Logout

*Insights at Your Fingertips - Just Upload and Ask*

User: [kmalhan@clarku.edu](mailto:kmalhan@clarku.edu)

Choose an option:

Clark University International Admissions ▼

Hello

Hello! How can I help you today?

Can you tell me about Clark University?

Here is some information about Clark University:

- **Founded:** 1887, making it the first graduate-only institution in the United States.
- **Reputation:** Known for rigorous academics, hands-on learning, and research focused on addressing global social and scientific challenges.
- **Location:** Worcester, Massachusetts, USA.
- **Student Body:** Attracts students from around the world, fostering a diverse and inclusive community.
- **Graduate School Highlights:**

Please type your question.



A close-up photograph of a person's hand, wearing a dark red long-sleeved shirt, reaching over a Go board. The board is a light-colored, textured surface with a grid of lines. It is covered with numerous black and white Go stones. Some stones are already placed on the board, while others are scattered around it. A blue circular container, likely for white stones, is visible in the background. A white plastic cup is on the right side of the board. The background is a soft, out-of-focus green, suggesting an outdoor setting. The lighting is warm and natural, creating a calm and focused atmosphere.

# Demo

Let's try Kanji





Thank You

Q & A