Live Guided Project - GenAI For Audio

INTERVIEW KICKSTART

Instructor: Emmanuel Awa

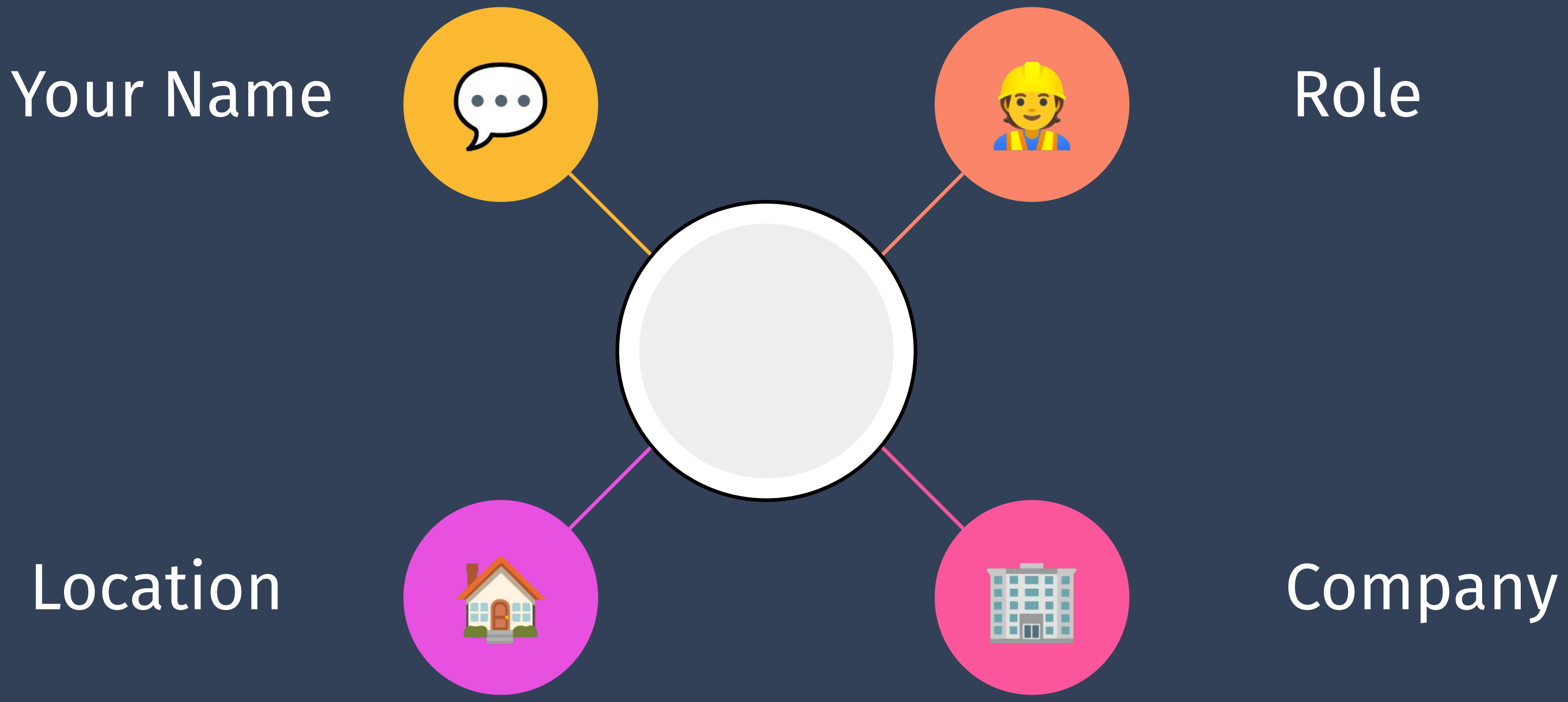
[Emmanuel Awa | LinkedIn]

Introduction – Emmanuel Awa

- ✔ In AI/ML space since 2016 with a focus on NLP
- ✔ School:
    - Undergrad: Physics with Electronics | Grad: Computer Science
- ✔ Post Graduate Work:
    - Amadeus for ~3 years
    - Microsoft for ~7.5 years
    - Google for a year now
- ✔ Career trajectory:
    - SWE => Big Data Engineer ==> AI Engineer ===> Data & Applied Scientist ====> Technical Solutions Arch - Applied GenAI

amadeus

Welcome to today's class, before we begin, pop into the chat:



Optimize Your Experience

✅ Interact with your instructors via Sunday live class.

✅ Don’t be shy to speak up and get clarifications !

✅ Solve your assignments, MCQs and other assessments and get feedback (Thursday review session)

✅ Use your resources! It’s your experience – what you put in is what you’ll get out.

How is Today's Content Relevant to My Job Role?

- **PMs:** Gain insights into customer interactions and enhance product strategies based on real-time data without needing to manually analyze each call.
- **TPMs:** Optimize call center workflows by automating repetitive tasks and improving lead engagement processes.
- **SDEs:** Develop and integrate AI models for natural language processing and speech-to-text systems within call center environments.
- **Engineering Managers:** Guide the deployment of AI-driven call center systems to enhance team performance and customer satisfaction.
- **DevOps Engineers:** Build and maintain the infrastructure required for scalable deployment and real-time processing of audio data.

Frequently Asked Question - 1

I don't have to code in my job role, how can I apply the knowledge from this course to my role?

You can use the knowledge acquired in this class to evaluate the feasibility of audio-related GenAI projects, estimate resource requirements and collaborate effectively with technical teams.

NOTE: This course does require some background in Python. If you have never coded in Python before, you may need to refer to external resources too. If you need help with the resources, please contact studentsupport@interviewkickstart.com

Frequently Asked Question - 2

What are some common pitfalls or challenges to avoid when implementing GenAI for audio projects?

- **Data quality issues:** Ensure clean, diverse, and representative audio datasets is crucial for training effective models.
- **Computational Resources:** GenAI for audio can be computationally intensive, however with careful planning allocation of adequate resources can be achieved.
- **Ethical Considerations:** Addressing biases, privacy concerns and copyright issues is essential in audio-related GenAI projects.

Today's Agenda

1. Generative AI for Audio Processing
2. Voice with LLMs
3. Architecture & Components
4. Putting it all together + Code

Generative AI For Audio Processing

Generative AI For Audio Processing

Motivation

Call centers often face challenges with handling high volumes of customer interactions, leading to inconsistent and inaccurate responses.

The goal of this project is to use Generative AI with, Audio modality, to improve efficiency and customer satisfaction by automating the analysis of recorded conversations and proving real-time engagement through a voice enabled call center bot.

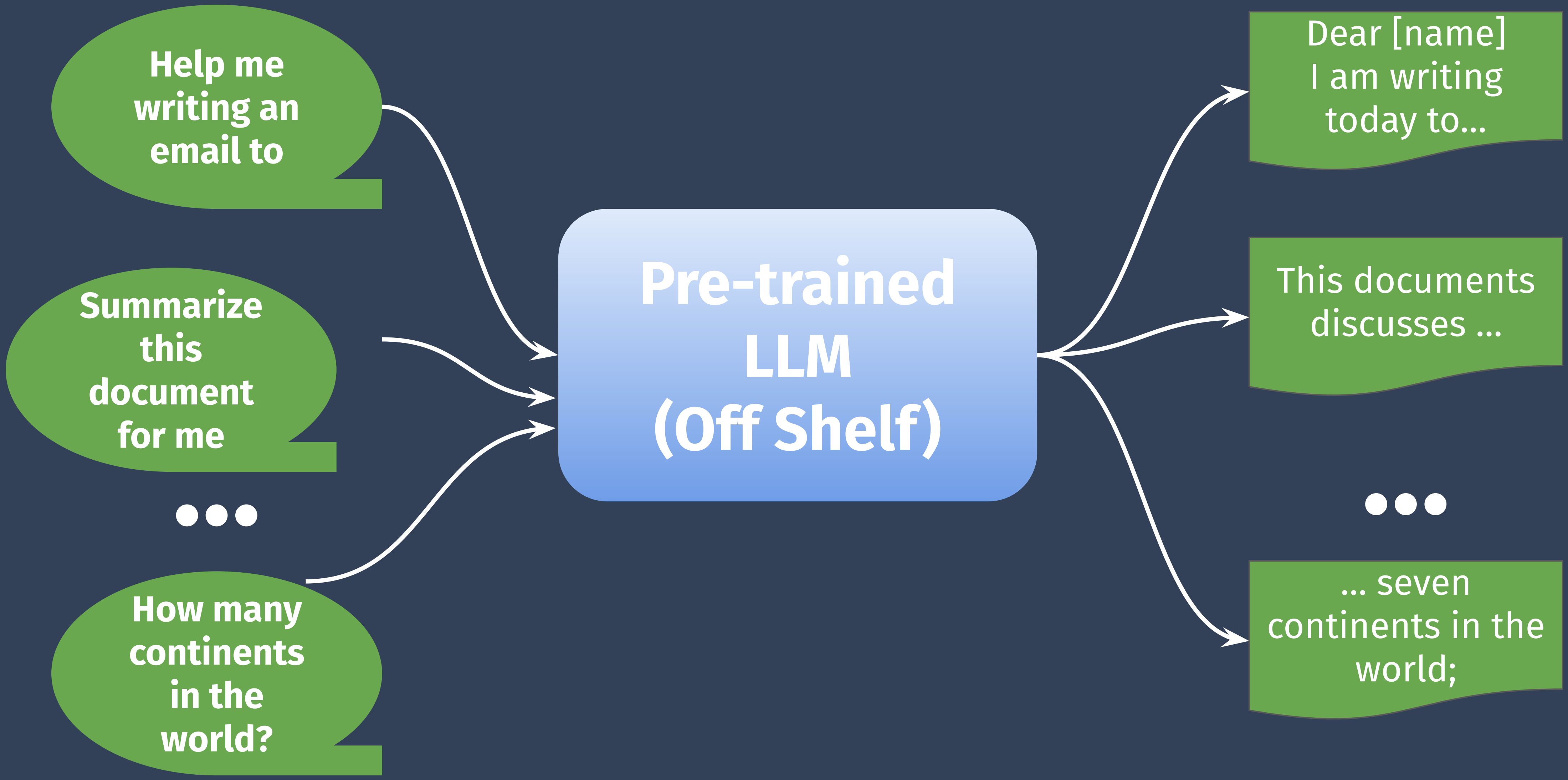
Generative AI For Audio Processing

Objectives

- Build a voice-activated bot capable of handling real-time conversations with leads in a call center environment
- Implement advanced speech processing techniques to identify, segment and label audio from multi-speaker conversations
- Utilize Generative AI models to generate accurate, context-aware responses for different customer scenarios
- Introduce the possibility of enhancing the call center experience by integrating the bot with data systems for better lead management and personalized service

Voice with LLMs

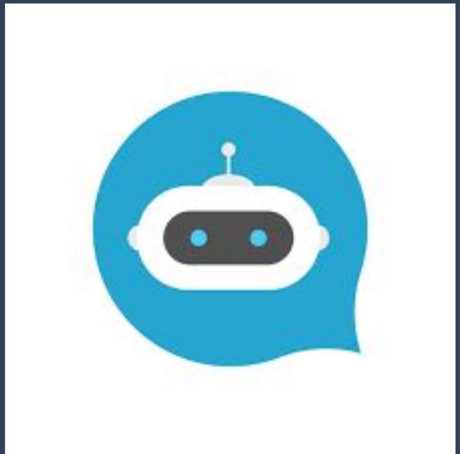
LLM


Help me writing an email to
Summarize this document for me
•••
How many continents in the world?
Pre-trained LLM (Off Shelf)
Dear [name] I am writing today to...
This documents discusses ...
•••
... seven continents in the world;

LLMs with Voice

LLMs with voice are a way for people to communicate effectively with a computer.

Having such capabilities can help improve the user experience for people to do their various daily tasks.

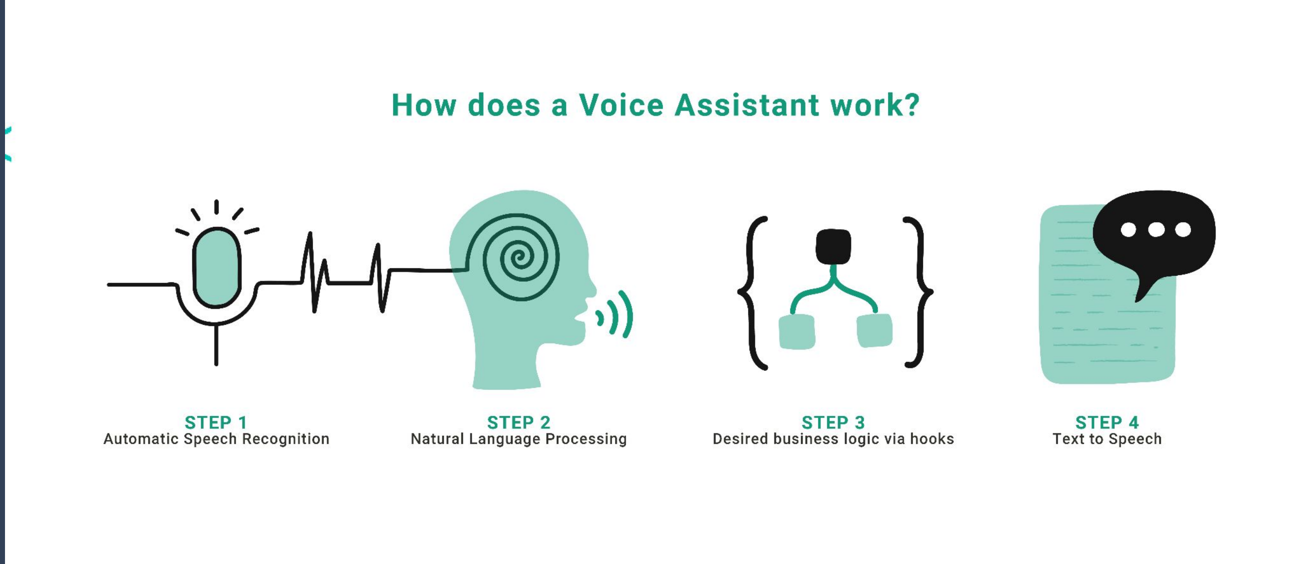


Key Components

1. Converting spoken language into text.
2. Processing the text to extract meaning and intent.
3. Generating a response based on the processed text and the LLM's training data.
4. Converting the generated text back into spoken language.

Applications of LLMs with Voice

Alexa
Siri
Google Voice Assistant



Before GenAI for Audio

Before GenAI, Siri had **Concatenative and Parametric Synthesis,** both had many limitations

Before GenAI for Audio

Concatenative Synthesis

Early on Siri and Google Voice would concatenate pre-recorded audio segments from a large database of speech fragments.

While this produced great audio, it didn't sound natural and lacked flexibility.

Before GenAI for Audio

Parametric Synthesis.

- Later, Siri and Google Voice utilized parametric synthesis models which were statistical models to generate speech.
- These systems would generate audio waveforms by adjusting parameters of a pre-defined model based on input text
- While this was an improvement over concatenative synthesis when it comes to flexibility, the voices would still sound robotic and less natural.

How were LLMs with voice an improvement?

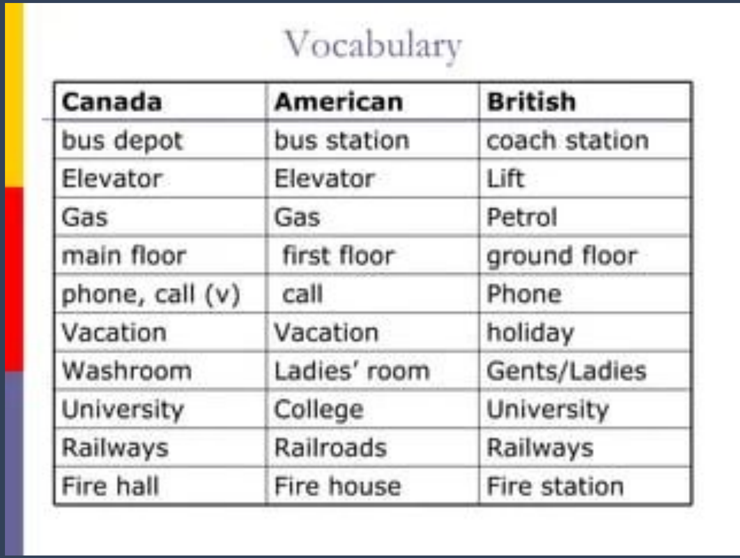
- Deep Learning Models that were autoregressive models that predicted audio samples based upon previous audio samples resulting in more natural sounding audio
- Transformers helped handle sequential data better by using self-attention mechanisms to produce more natural and coherent speech
- Transformers captured the bigger picture of long-term relationships within a piece of audio better than traditional models such as LSTM.
- Transformers were much faster at processing information quickly than older models.

Limitations

- Can not understand different accents or dialects or different types of voices
- Do not work well if a person asks the chat bot in a busy street with lots of noise
- Sometimes don't understand the context of the user's question

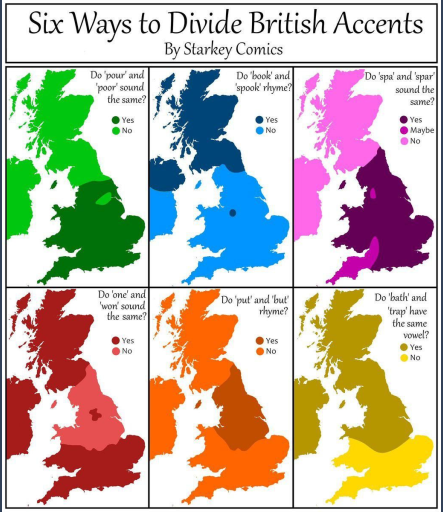
How to overcome these limitations?

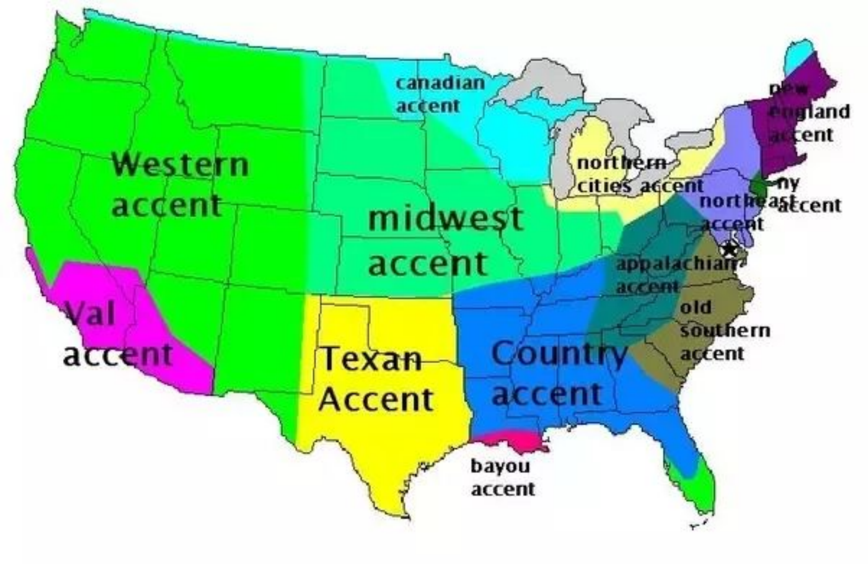
- Wider range of accents, dialects, and languages are used in the training datasets for LLMs with voice
- Noise cancellation and background noise suppression technologies
- Using GPUs for Machine Learning to improve the time it takes the LLM to run and process what is said



Vocabulary

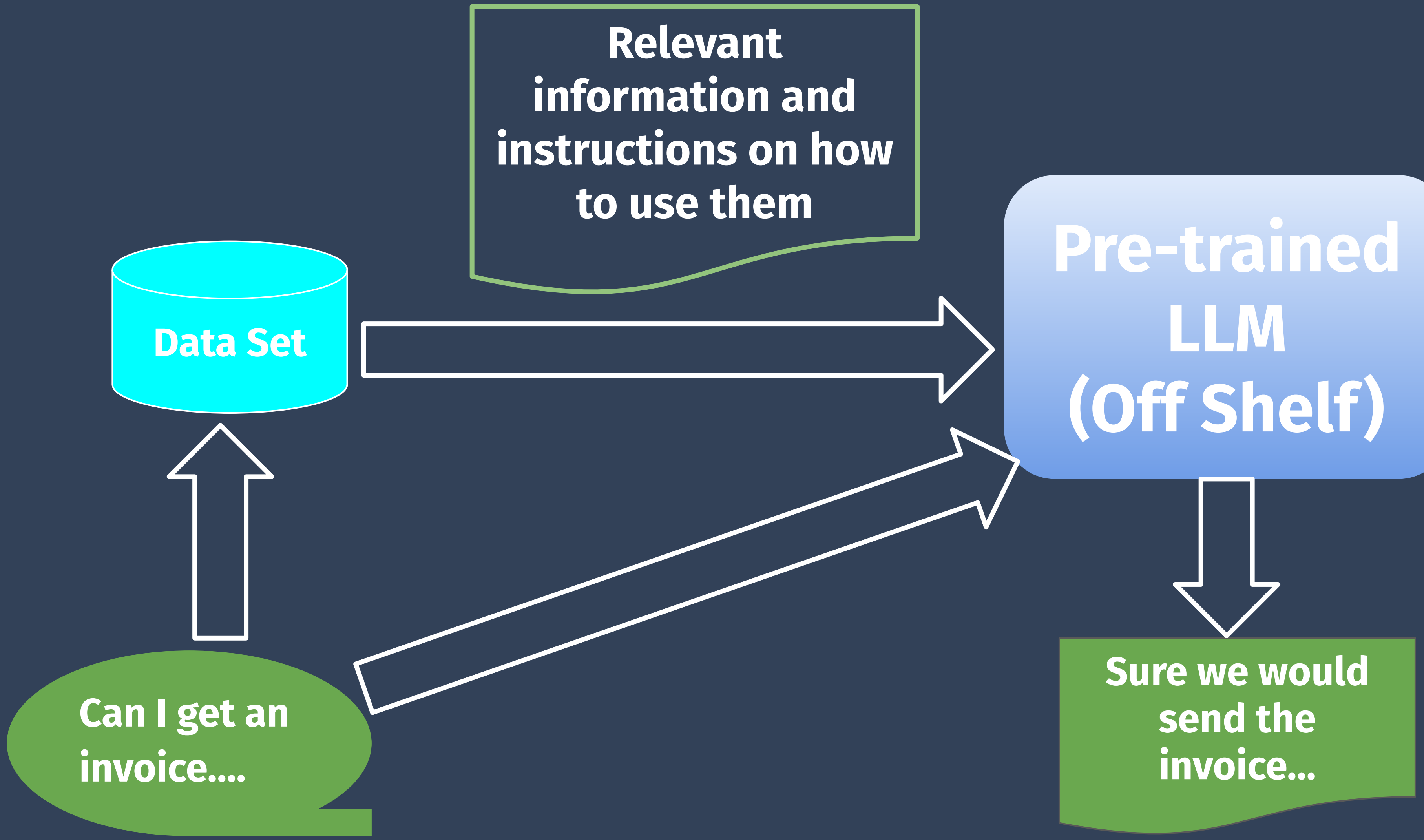
Canada	American	British
bus depot	bus station	coach station
Elevator	Elevator	Lift
Gas	Gas	Petrol
main floor	first floor	ground floor
phone, call (v)	call	Phone
Vacation	Vacation	holiday
Washroom	Ladies' room	Gents/Ladies
University	College	University
Railways	Railroads	Railways
Fire hall	Fire house	Fire station



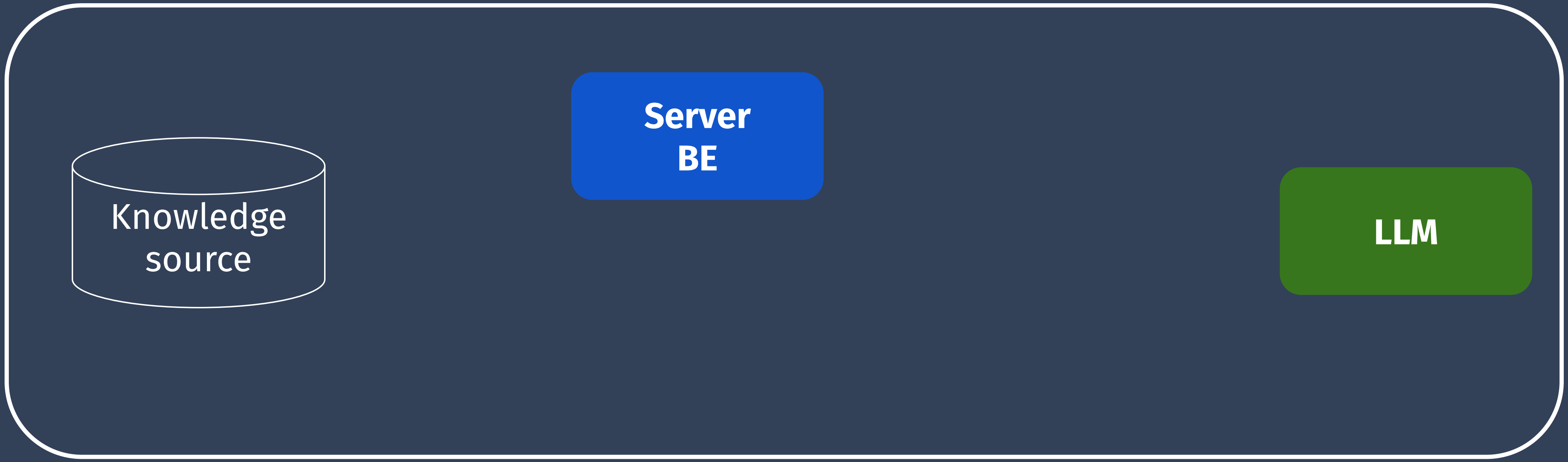


Architecture
and
Project Components

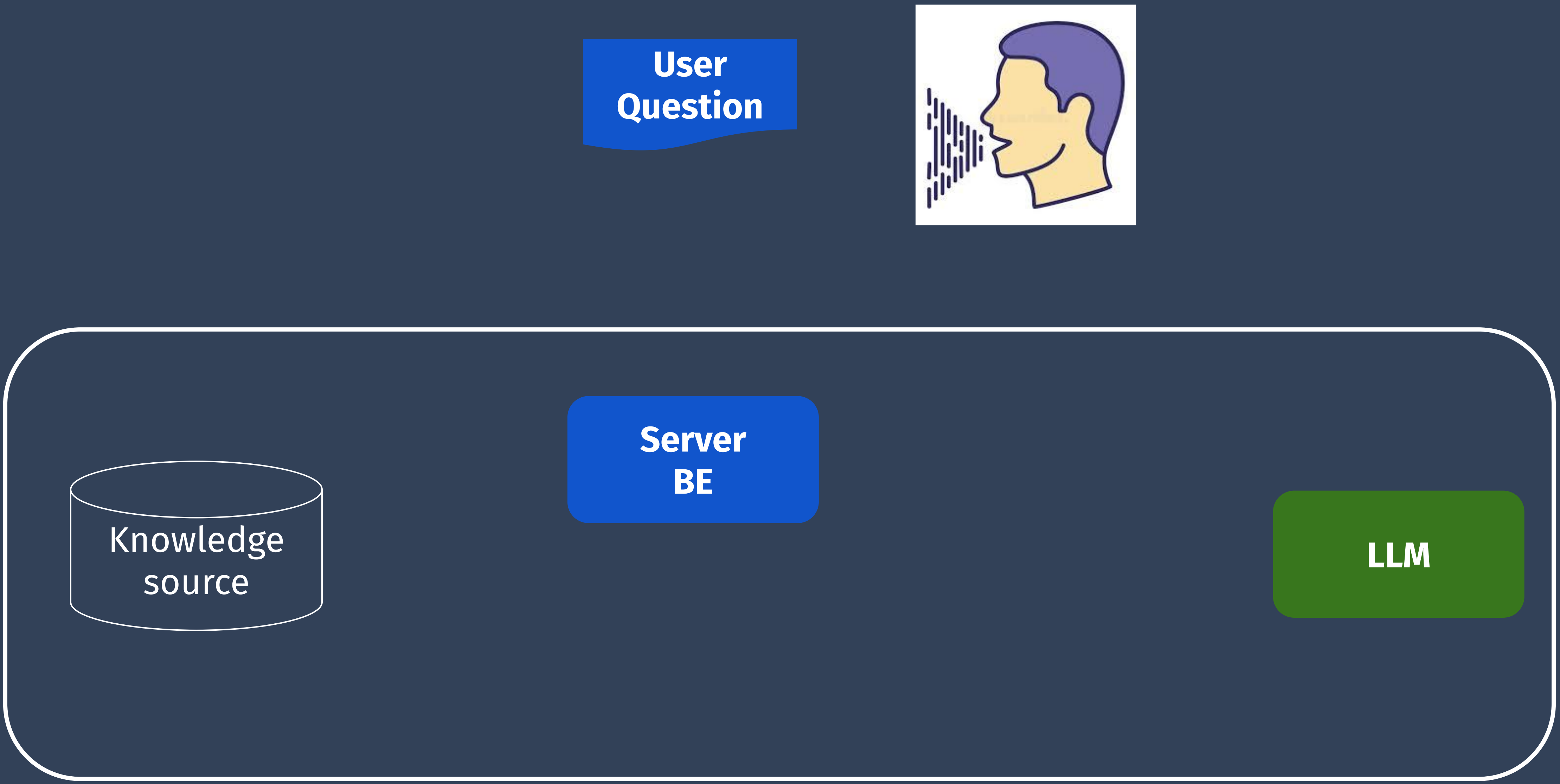
How Does it Work?



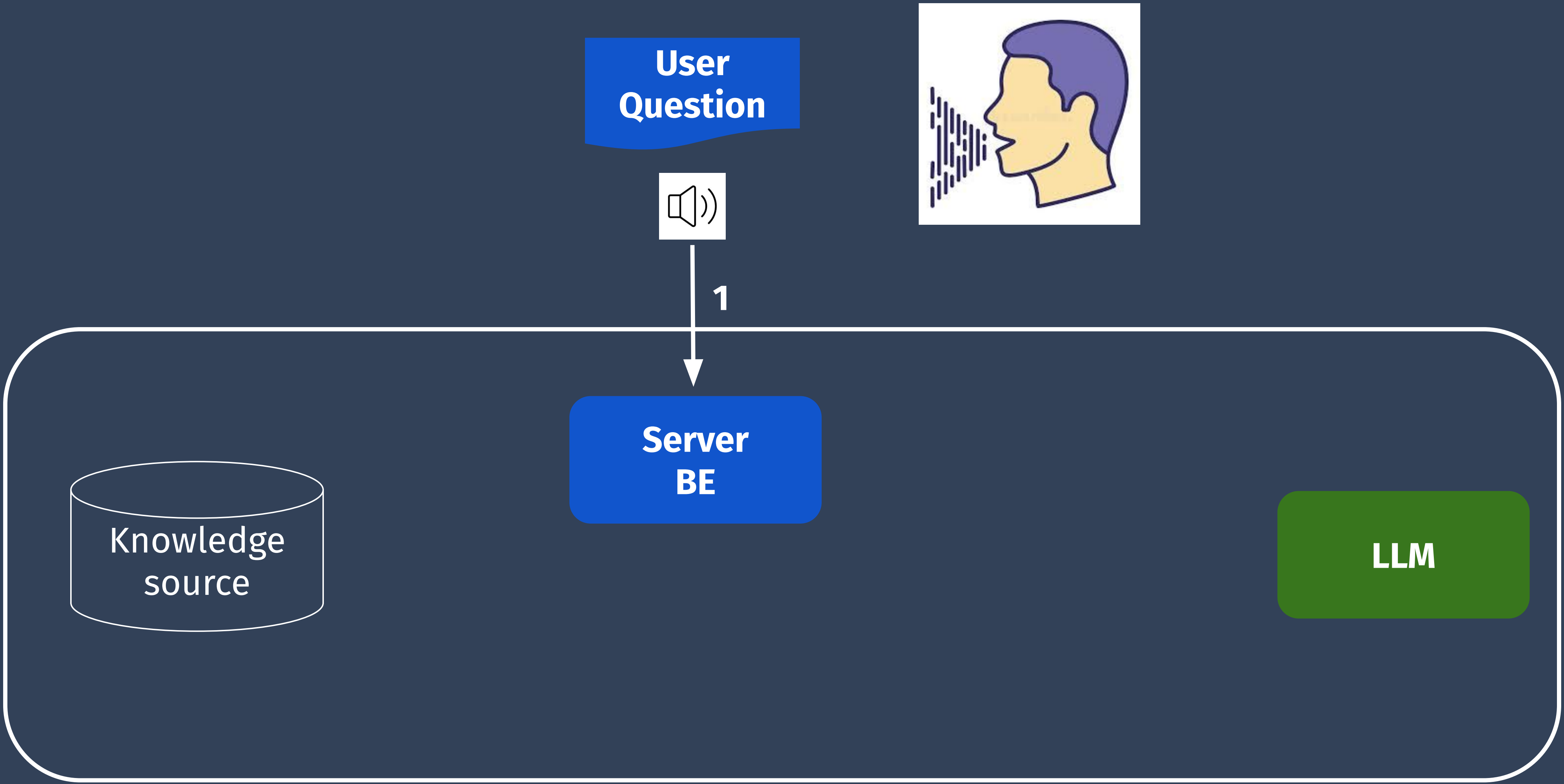
Architecture



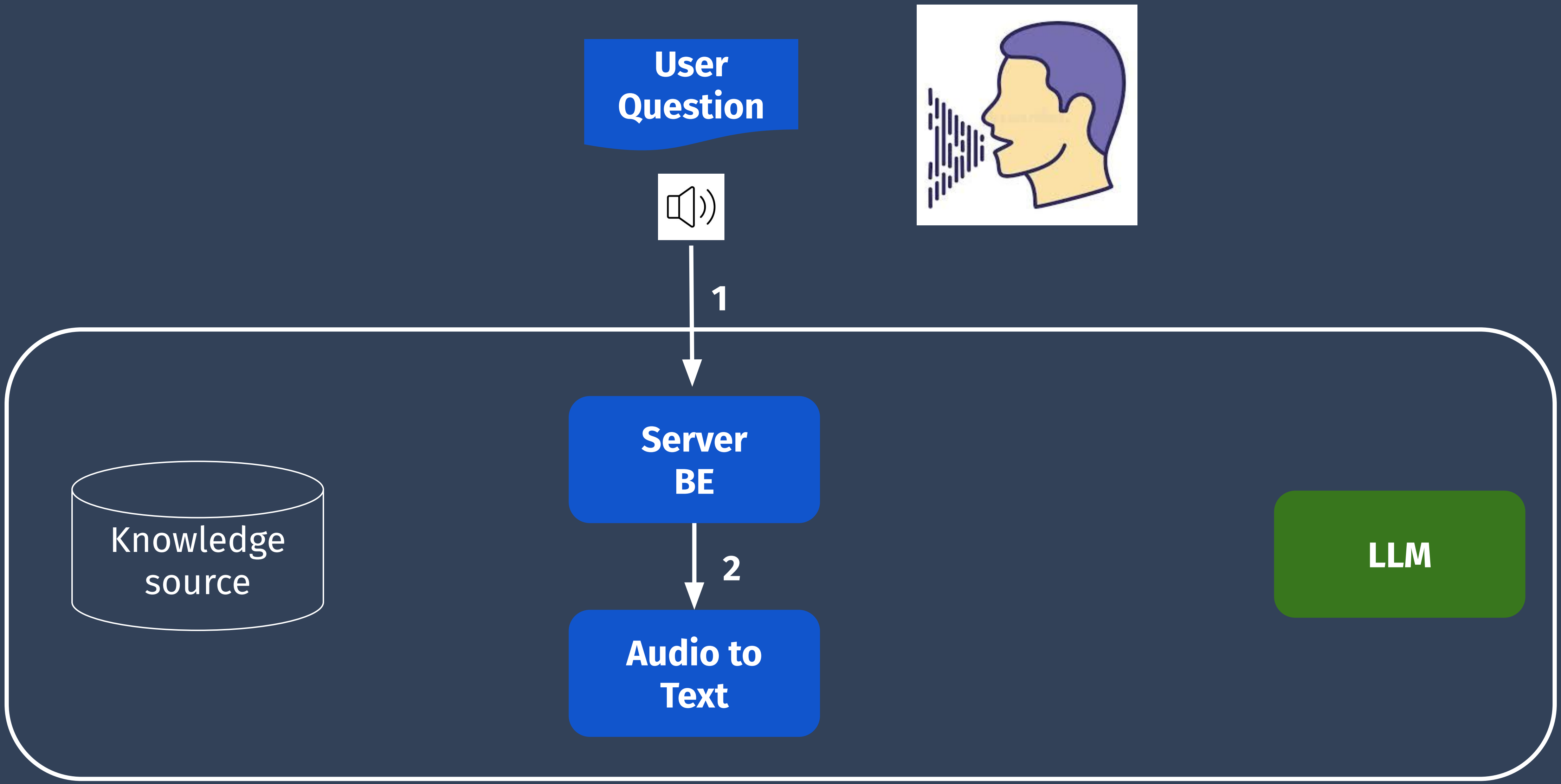
Architecture



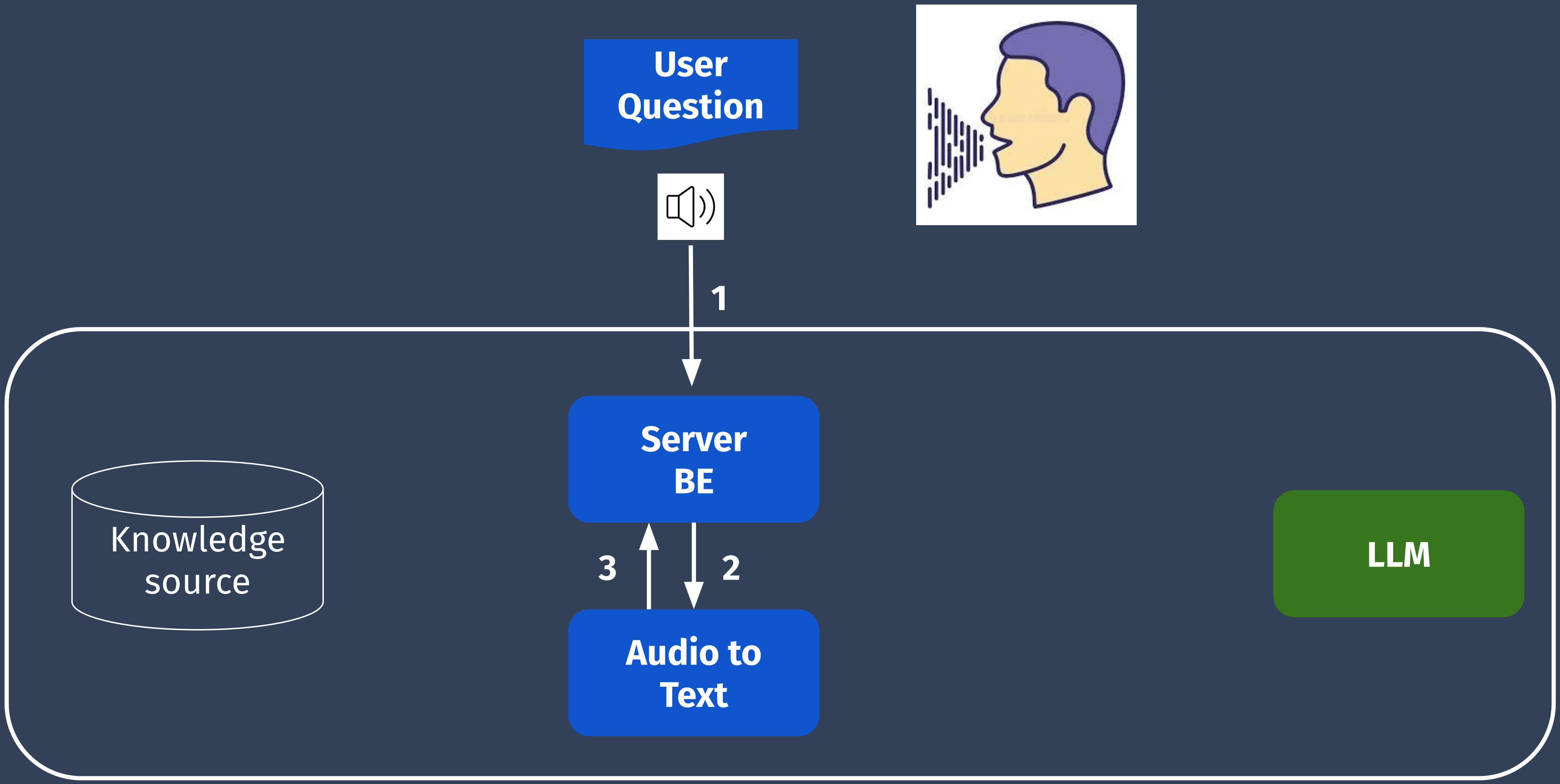
Architecture



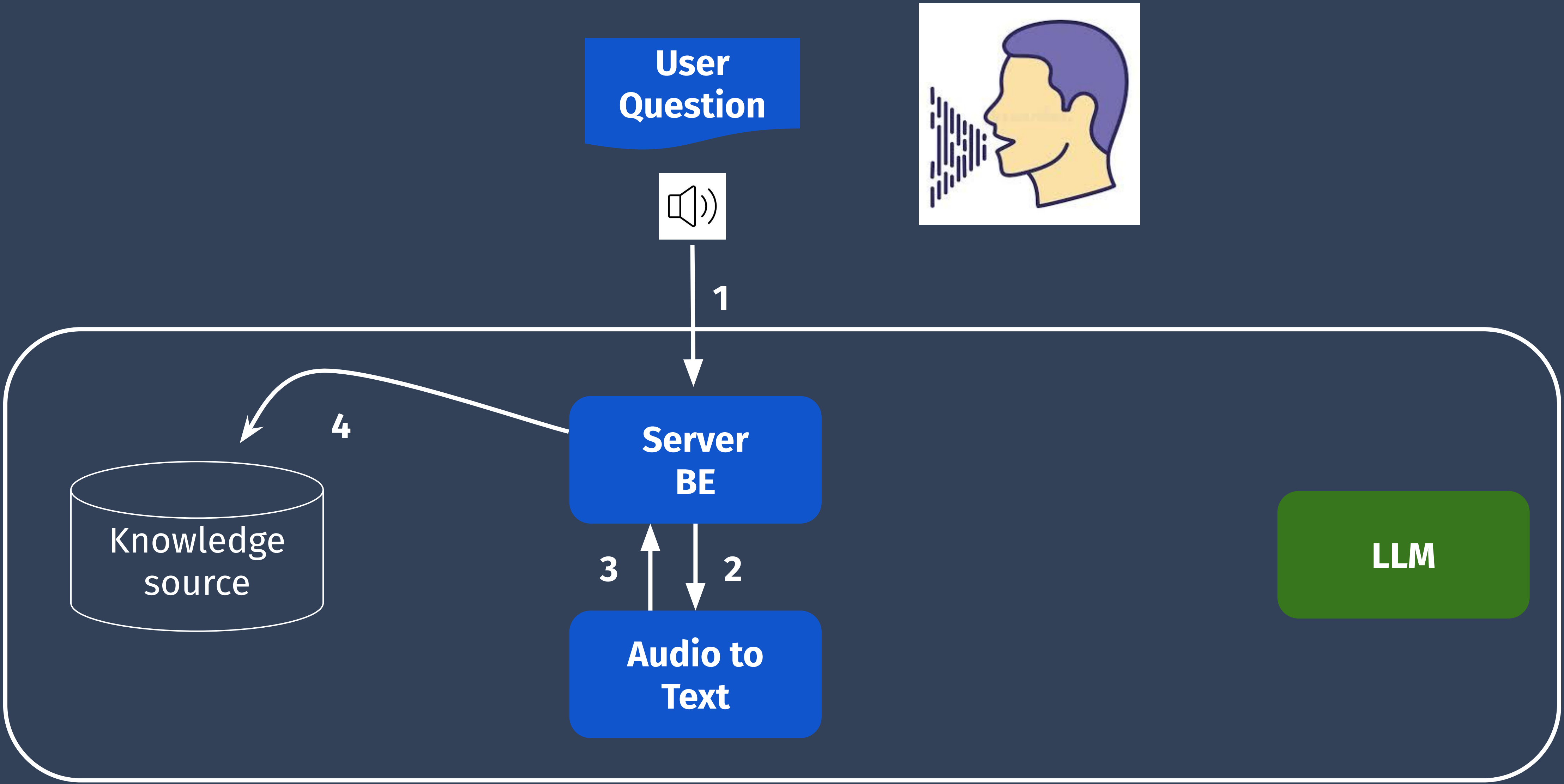
Architecture



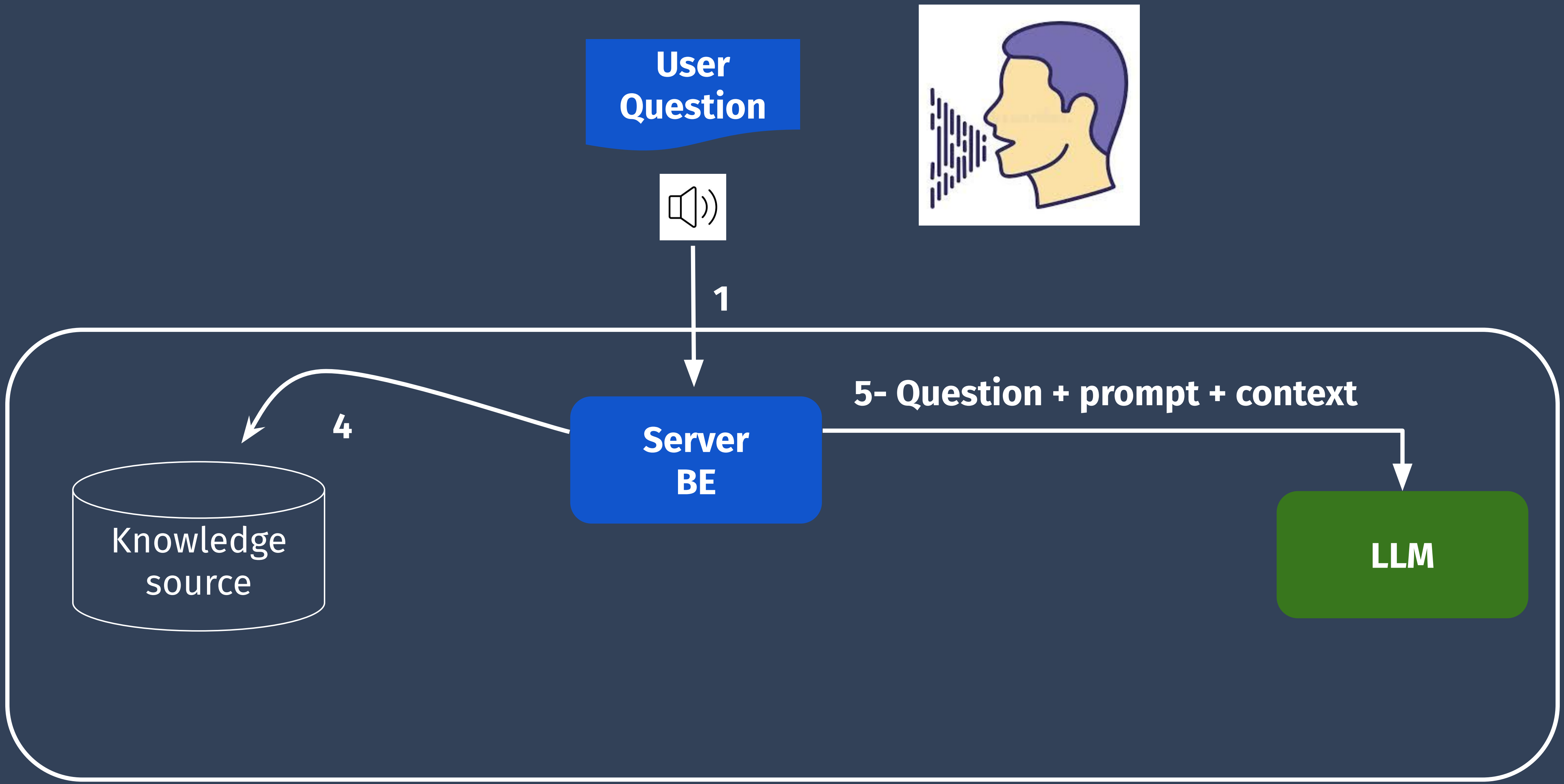
Architecture



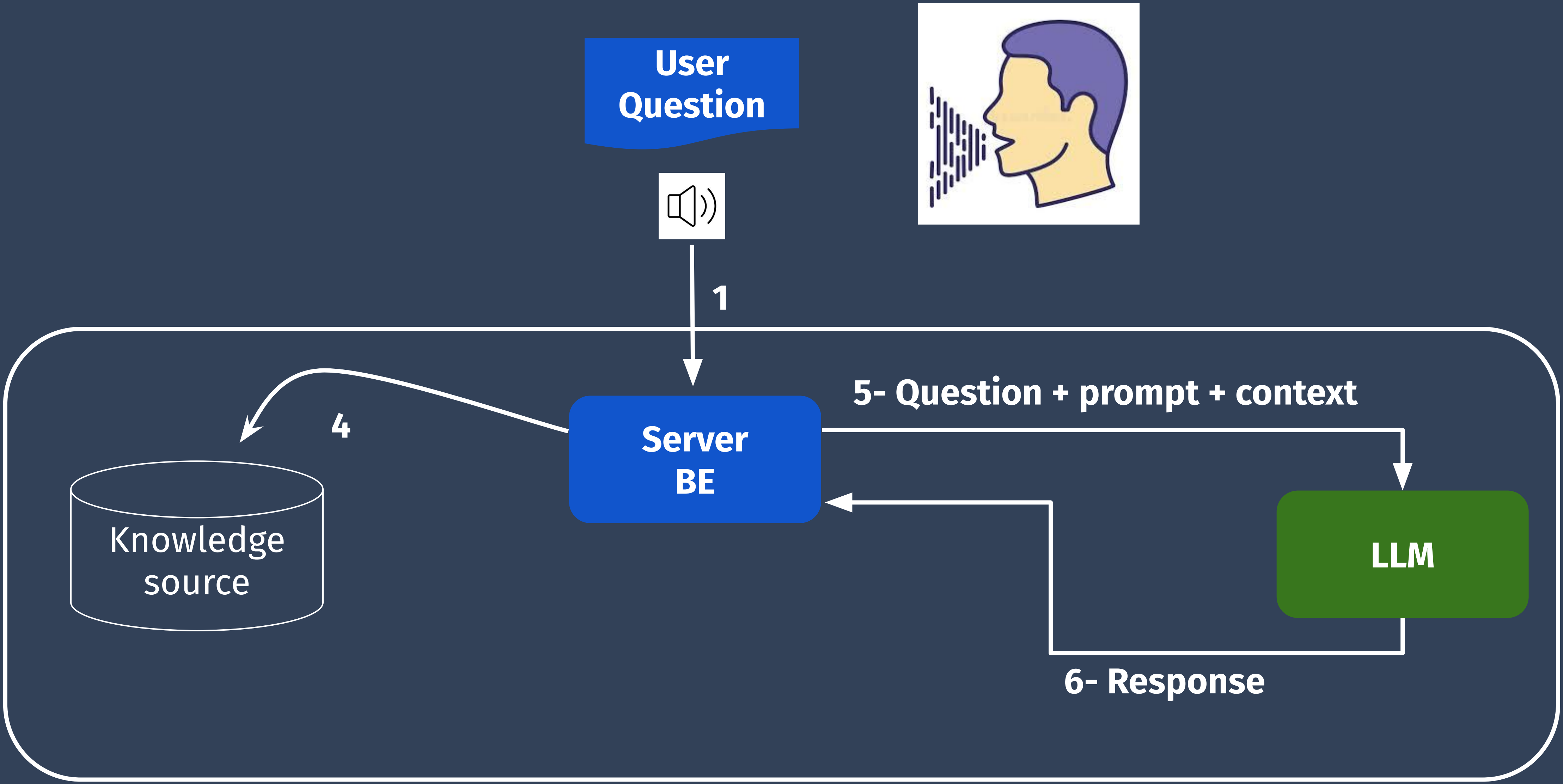
Architecture



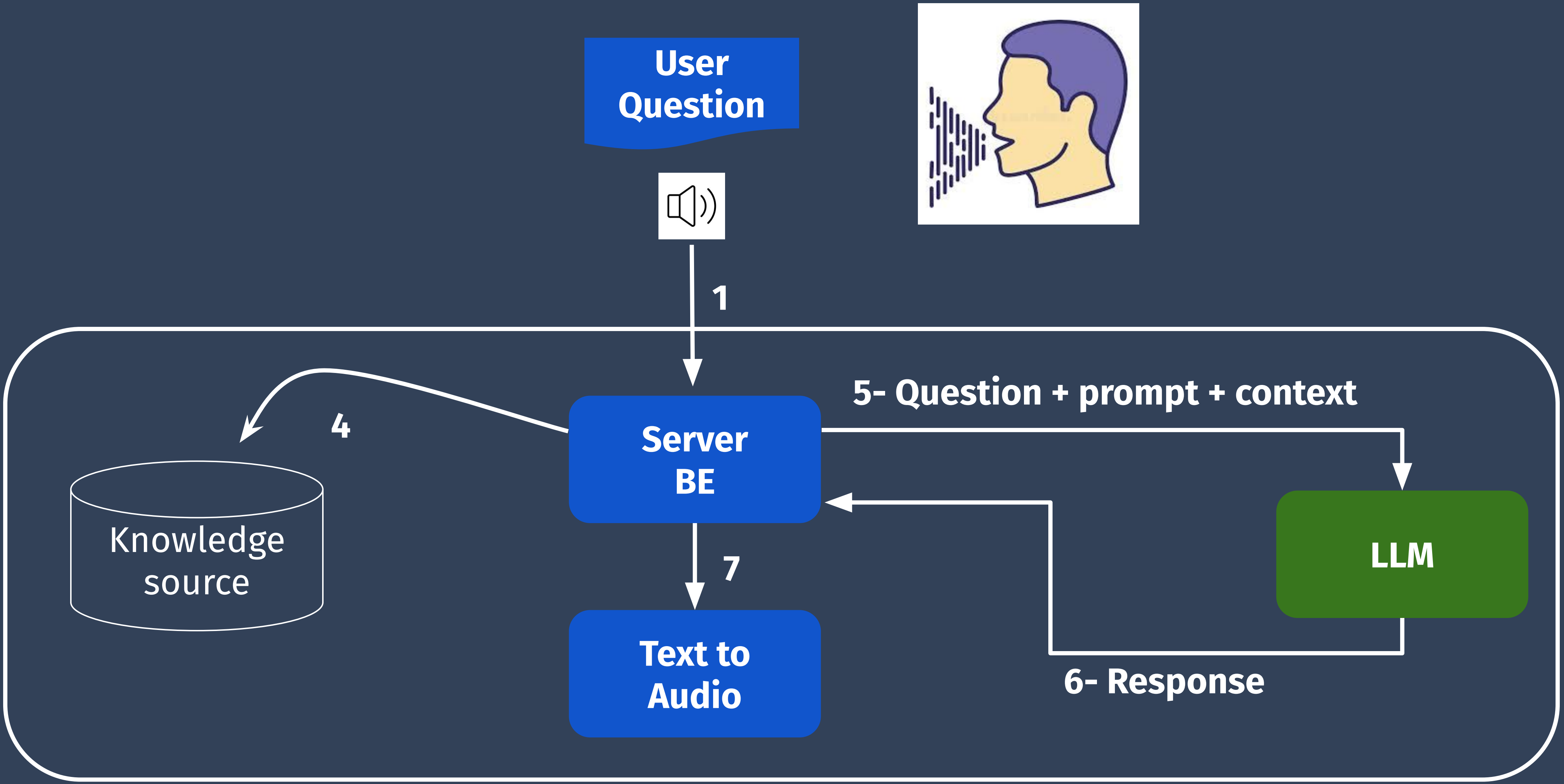
Architecture



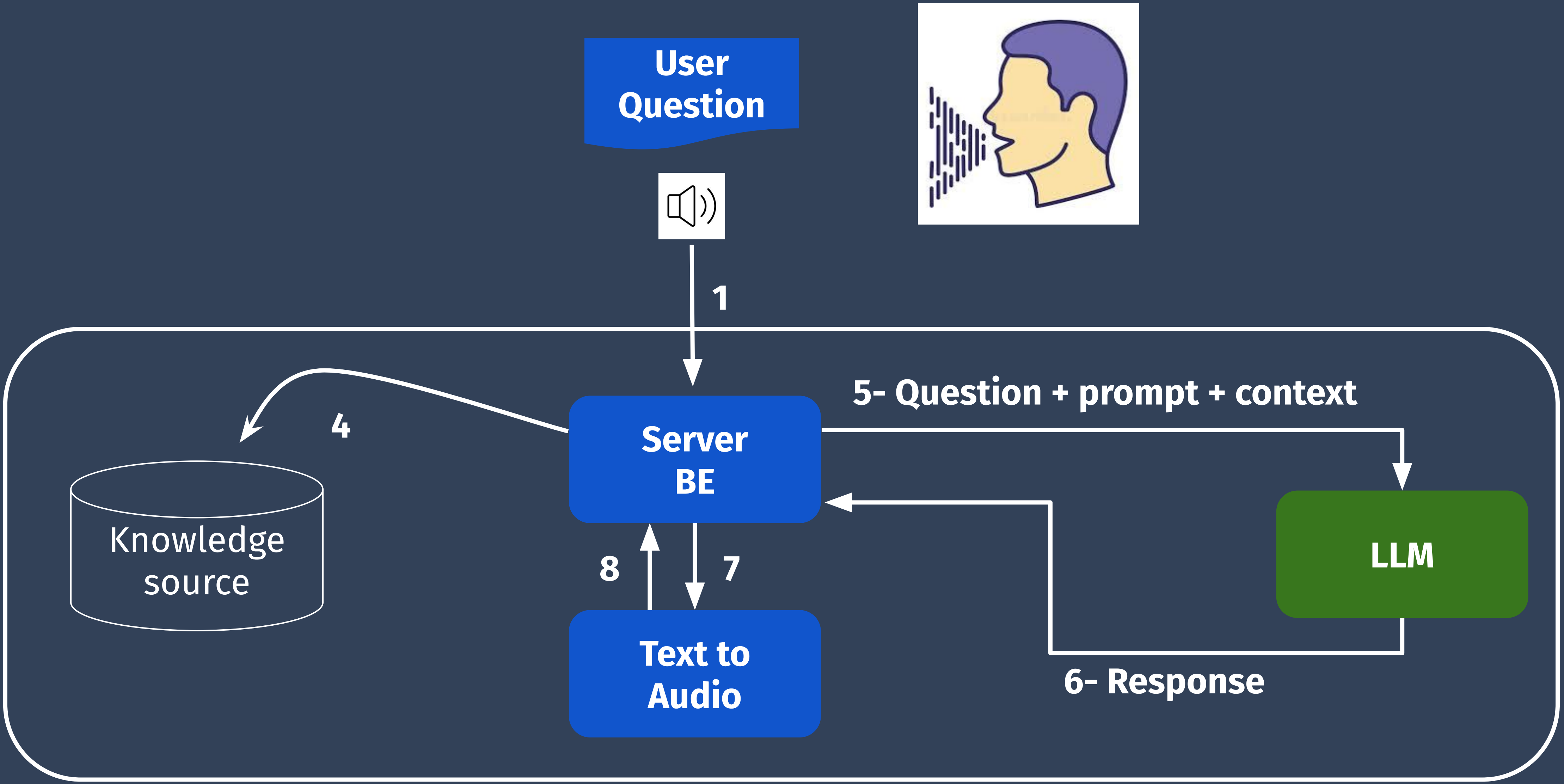
Architecture



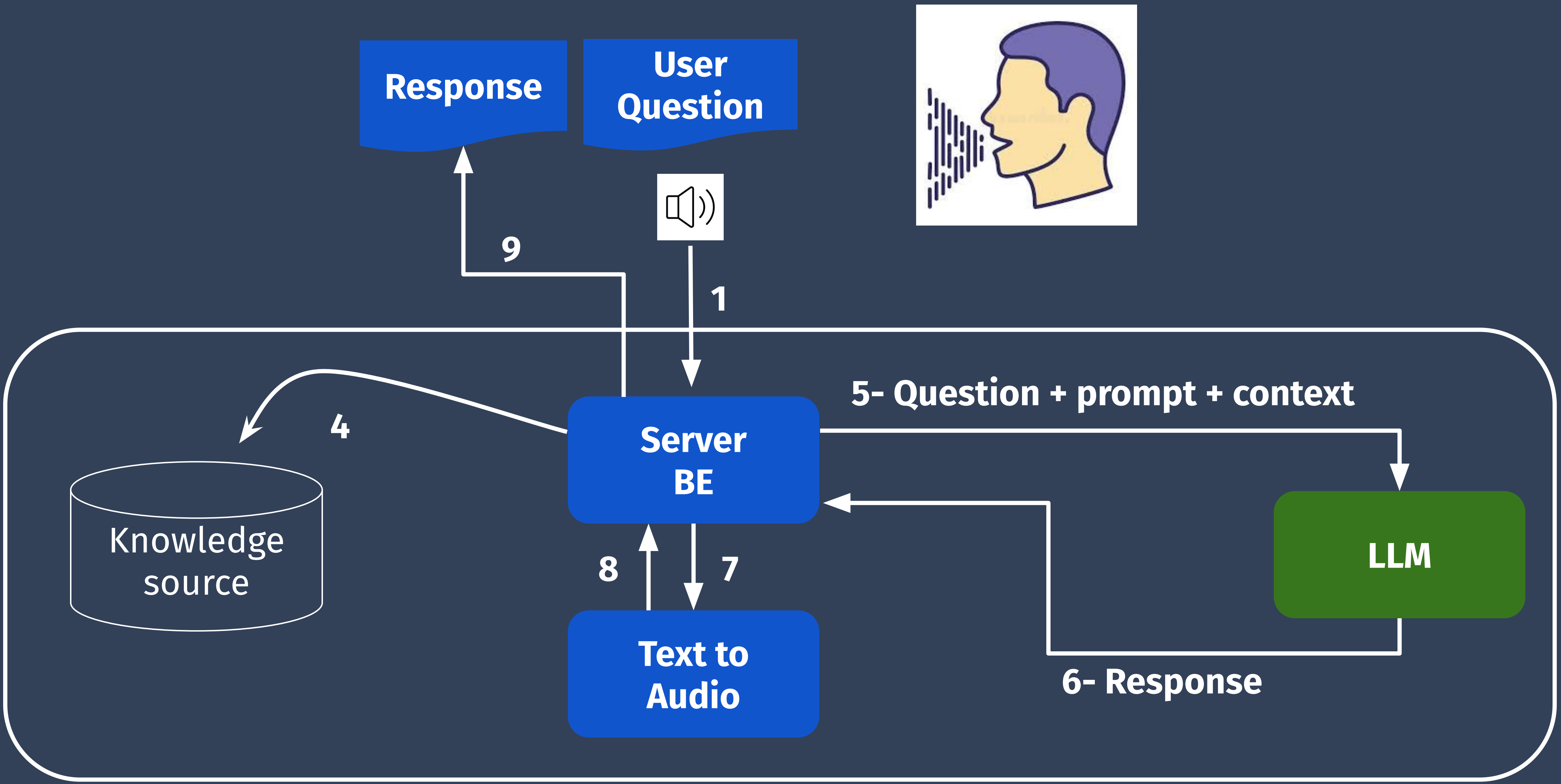
Architecture



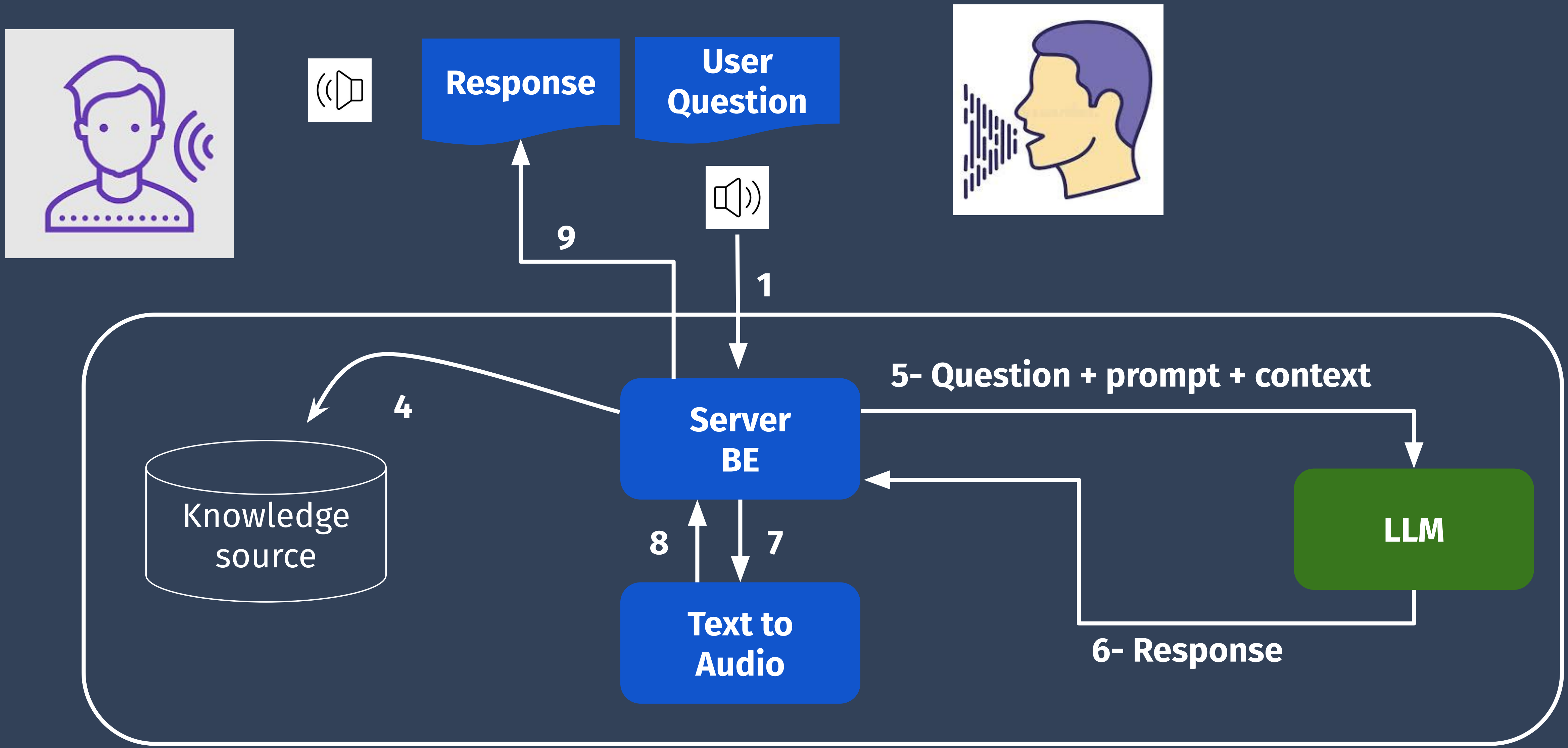
Architecture



Architecture



Architecture



Putting it all together

Audio Input Processing

- Use OpenAI Whisper for automatic speech recognition
- Implement Voice Activity Detection (VAD) to segment the audio

Text Processing and Contextual Understanding

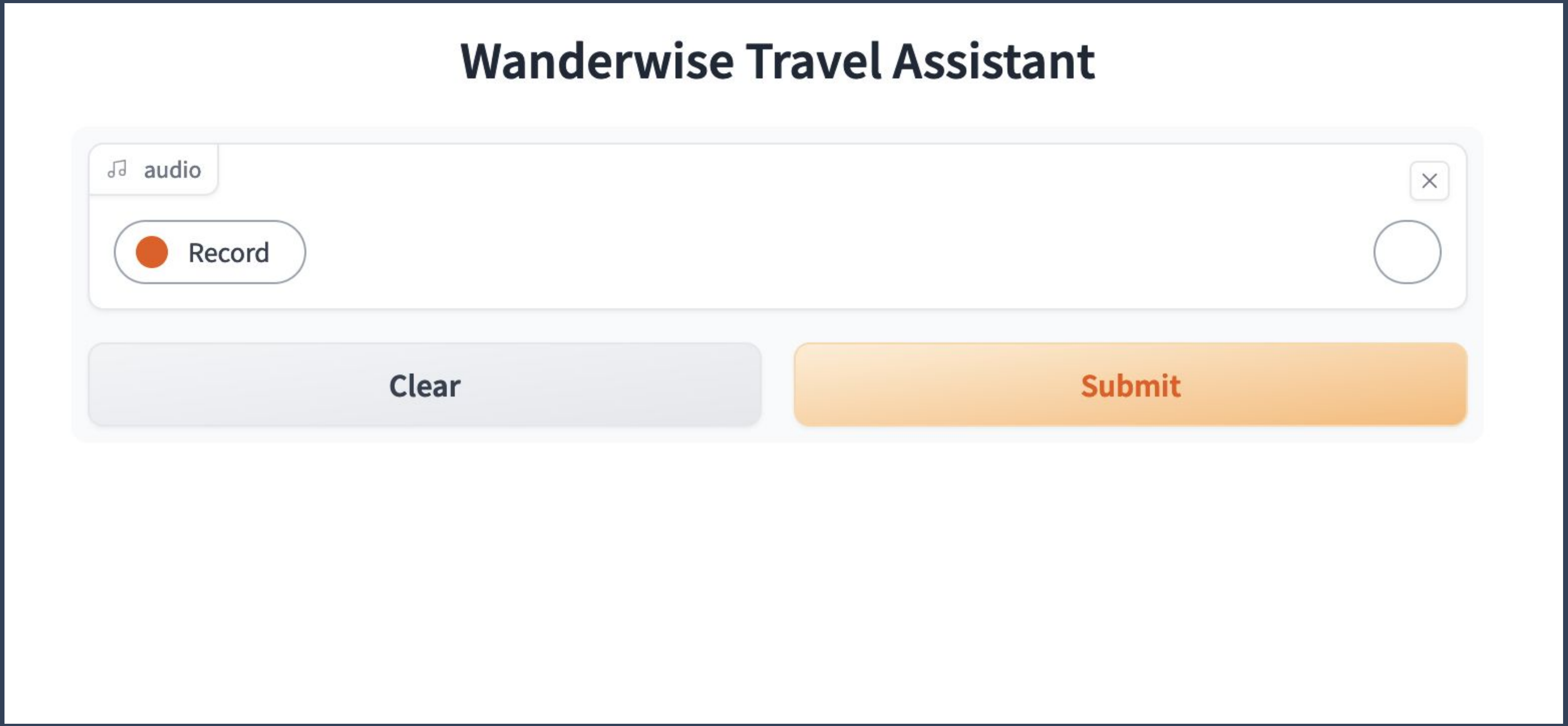
- Scrape data using LangChain toolkit if needed
- Extract relevant features from the audio
- Convert audio segments into text for the OpenAI GPT model
- Use LangChain toolkit for data preprocessing
- Implement Voice Activity Detection (VAD) to segment the audio

Real-time Conversational Request and Responses

- Use FFmpeg to handle and process incoming audio files
- Utilize OpenAI GPT model to generate context-aware responses
- Convert text back to audio using a text-to-speech module (e.g. pyttsx3)

Application UI

- Use Gradio to build an interactive application



30,000 Overview Of Project Structure

app

- Contains the main application files

data

- Stores various types of data for the chatbot

genai_voice

- Core project code that is installable - bots, config, models, processing...

Libs

- Windows version of ffmpeg

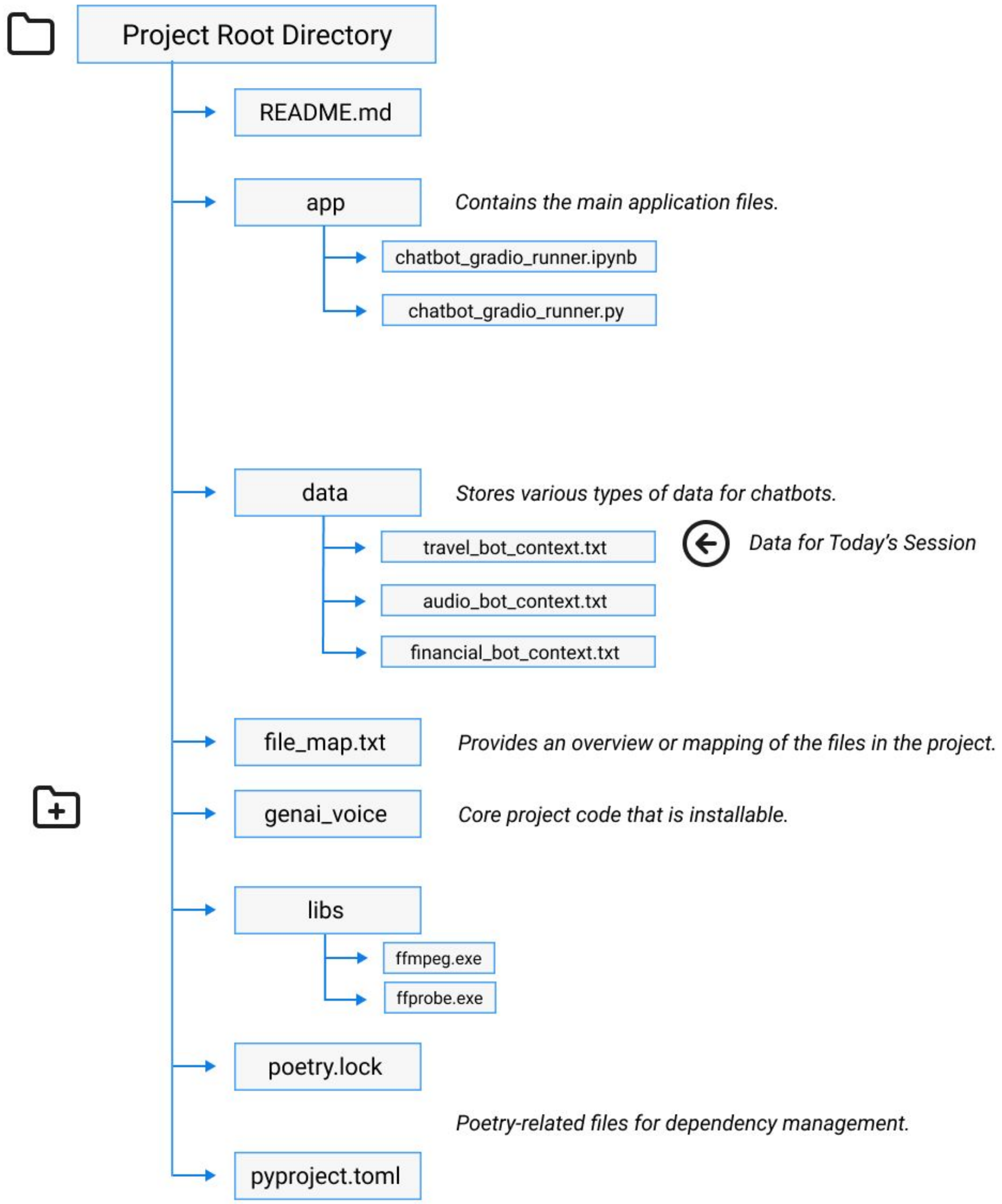
pyproject.toml/poetry.lock

- Poetry-related files for dependency management

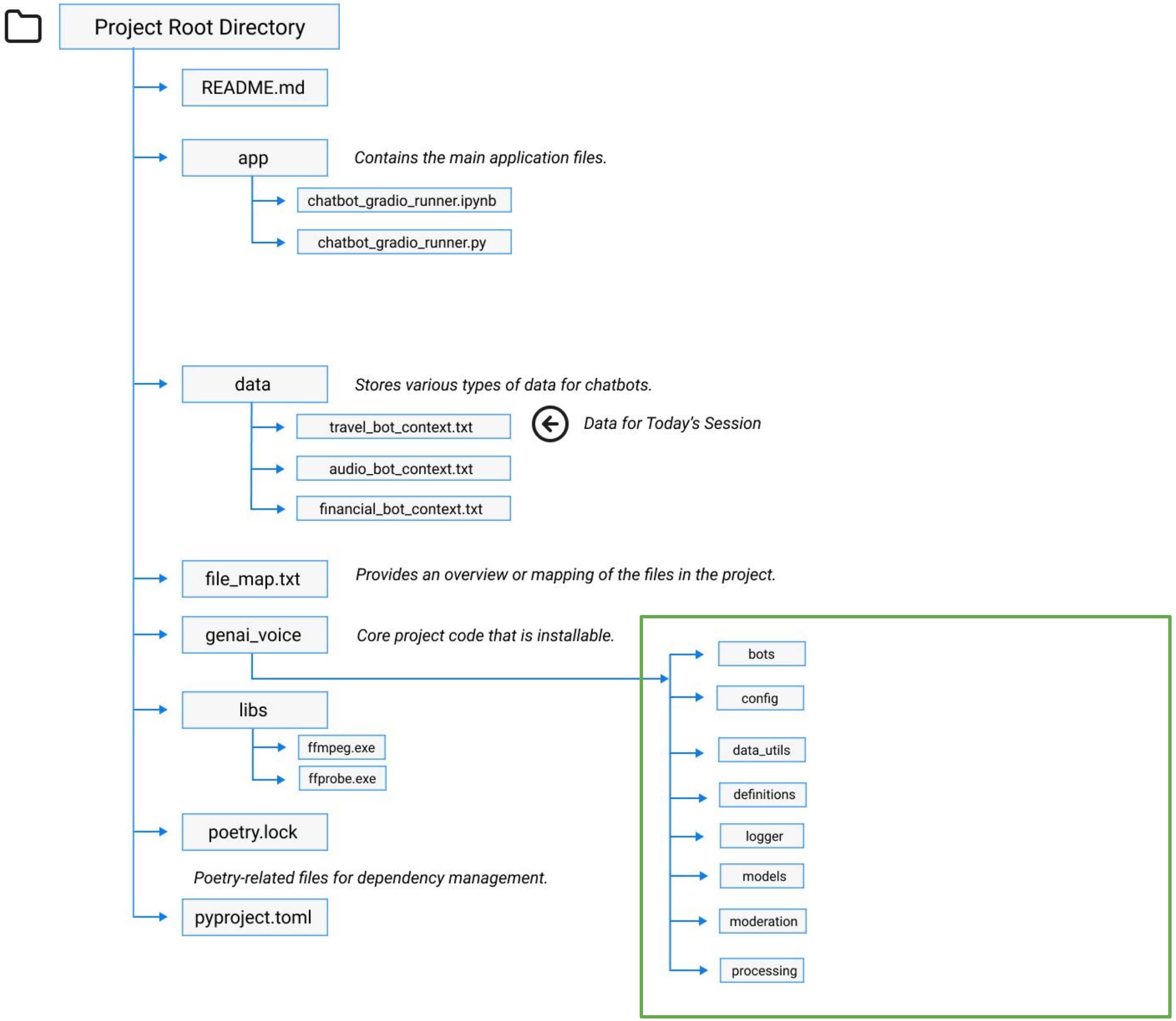
Code Directory

–

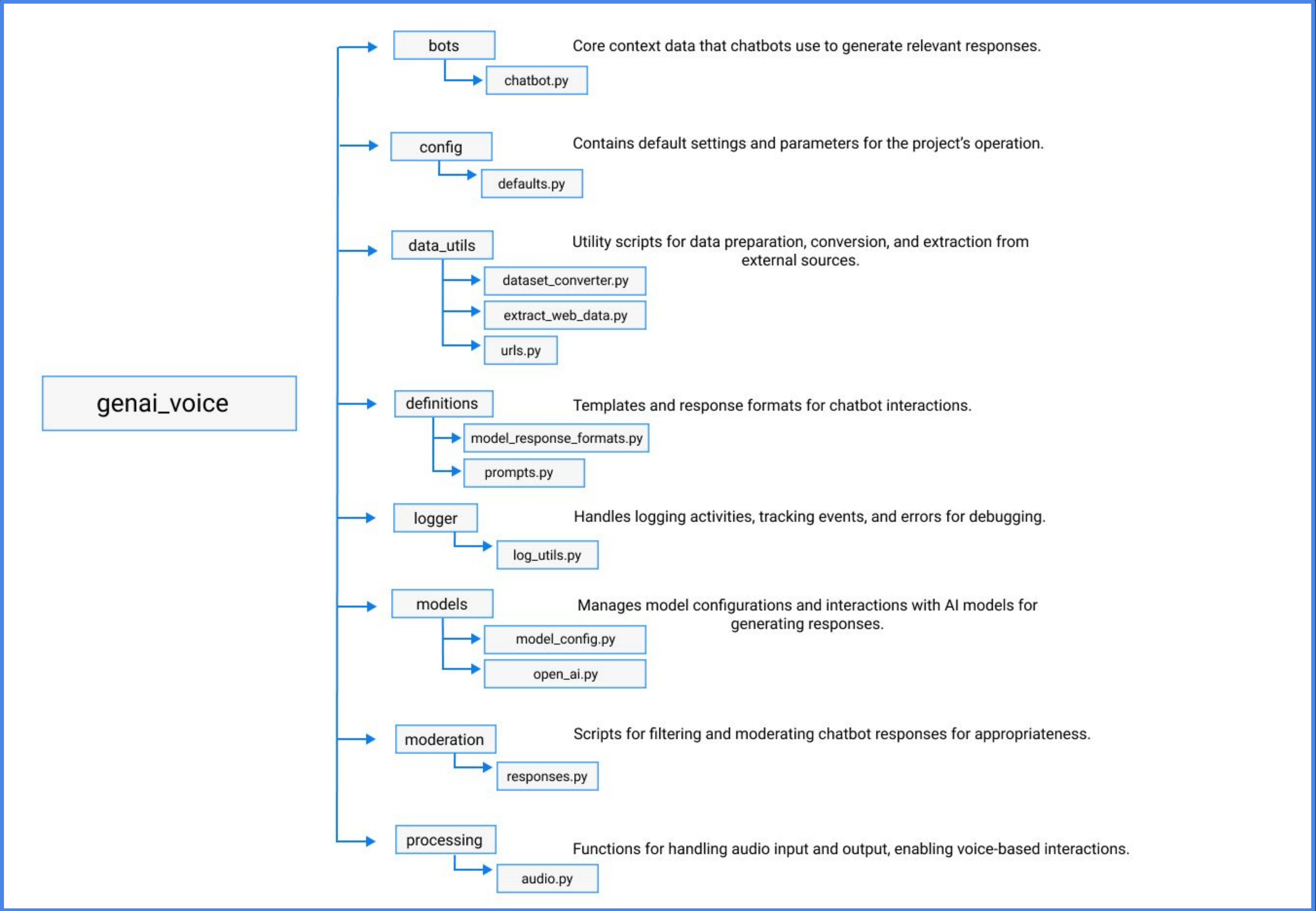
Deep Dive



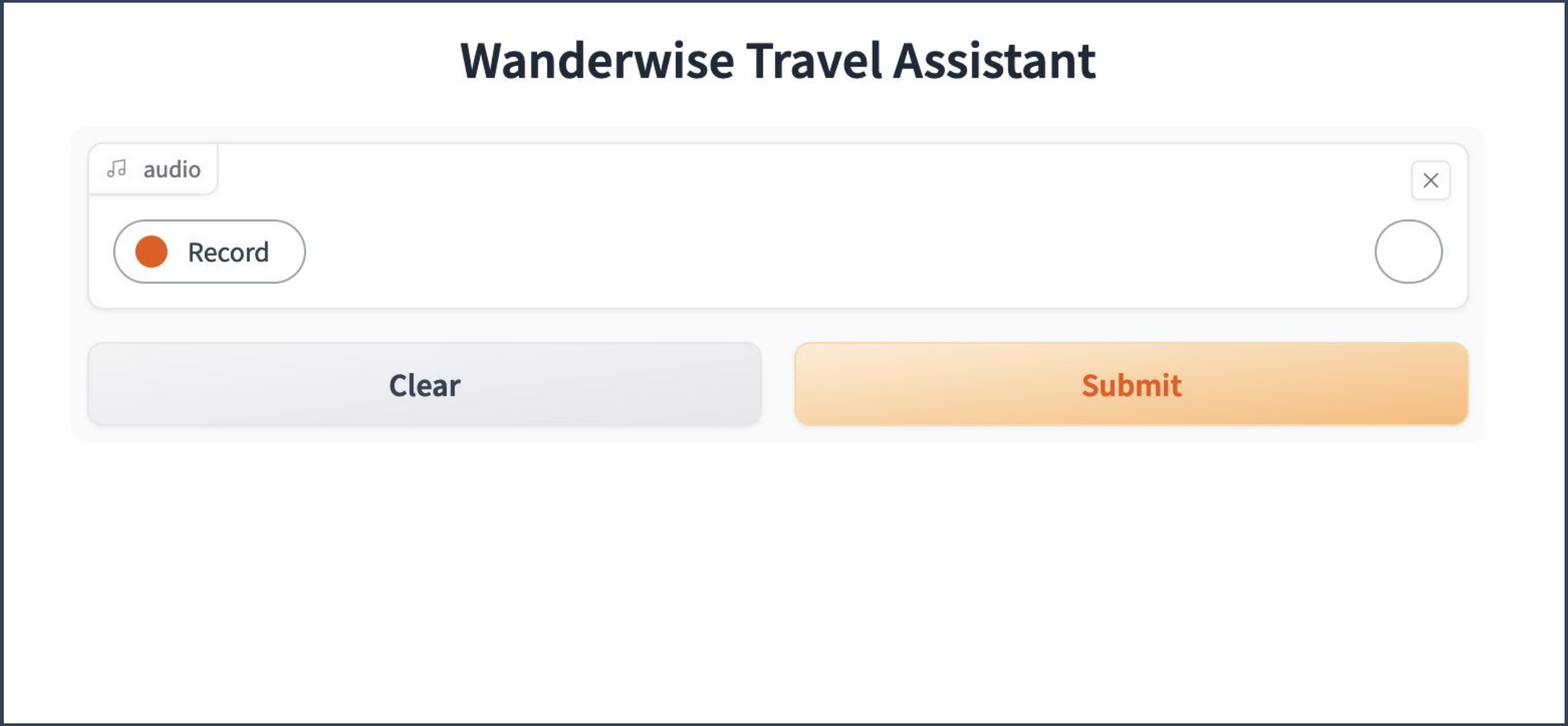
Project Structure



Project Structure


genai_voice
bots
chatbot.py
Core context data that chatbots use to generate relevant responses.
config
defaults.py
Contains default settings and parameters for the project's operation.
data_utils
dataset_converter.py
extract_web_data.py
urls.py
Utility scripts for data preparation, conversion, and extraction from external sources.
definitions
model_response_formats.py
prompts.py
Templates and response formats for chatbot interactions.
logger
log_utils.py
Handles logging activities, tracking events, and errors for debugging.
models
model_config.py
open_ai.py
Manages model configurations and interactions with AI models for generating responses.
moderation
responses.py
Scripts for filtering and moderating chatbot responses for appropriateness.
processing
audio.py
Functions for handling audio input and output, enabling voice-based interactions.

Final App



Going To Code

Summary

01 Generative AI For Audio Processing

02 Voice with LLMs

03 Architecture and Components of Project

04 Putting it all Together/Code Structure

05 Hands-on Exercises and Practical Applications

Next Steps

- Revise the concepts learnt in today's session through class recording and post-class videos

- Register for Technical Coaching Session through Xpert Connect option on Uplevel

- Consume the Post Class Content on GenAI for Audio.

What’s Coming Next?

Now that you have grasped how to implement a voice-based chatbot using advanced AI models, **we will move next to the Domain Specific Sessions.**


Thank You