Term Project Final Presentation.

Topic: VoiceChain: Secure Automated IVR with LLM Integration.

Sushovan Pan, ID: B2330054

Guide: Champak Dutta.

RKMVERI

TOPIC

- Overview 3
- Objective 4
- Key features 5
- System Architecture 6-9
- Tech stack overview 10-12
- Experiments & Results 13-15
- **Future work 16-18**
- Use Case 19
- Benifits 20
- Conclusion 21

Overview

VoiceChain is a system that integrates automated Interactive Voice Response (IVR) with a Large Language Model (LLM) and supports blockchain-based call record storage as a future enhancement.

The system takes in user audio queries, processes them into text using speech recognition, and then passes the text to an LLM for intelligent query resolution.

Additionally, VoiceChain planned to use blockchain technology to securely store call records, ensuring tamper-proof storage as a future enhancement.

Objectives

1. To develop a highly secure, scalable, and efficient IVR system that can intelligently process user audio queries.
2. Integration of **LLM** for handling complex, real-time query responses.
3. Implementing **emotion recognition** for analyzing the emotional tone of the user enhances the system's response.
4. Seamless conversion between speech and text for natural communication with the system.

Key Features

Audio Query Handling: Users can interact with the system via VoIP or softphone by submitting audio queries.

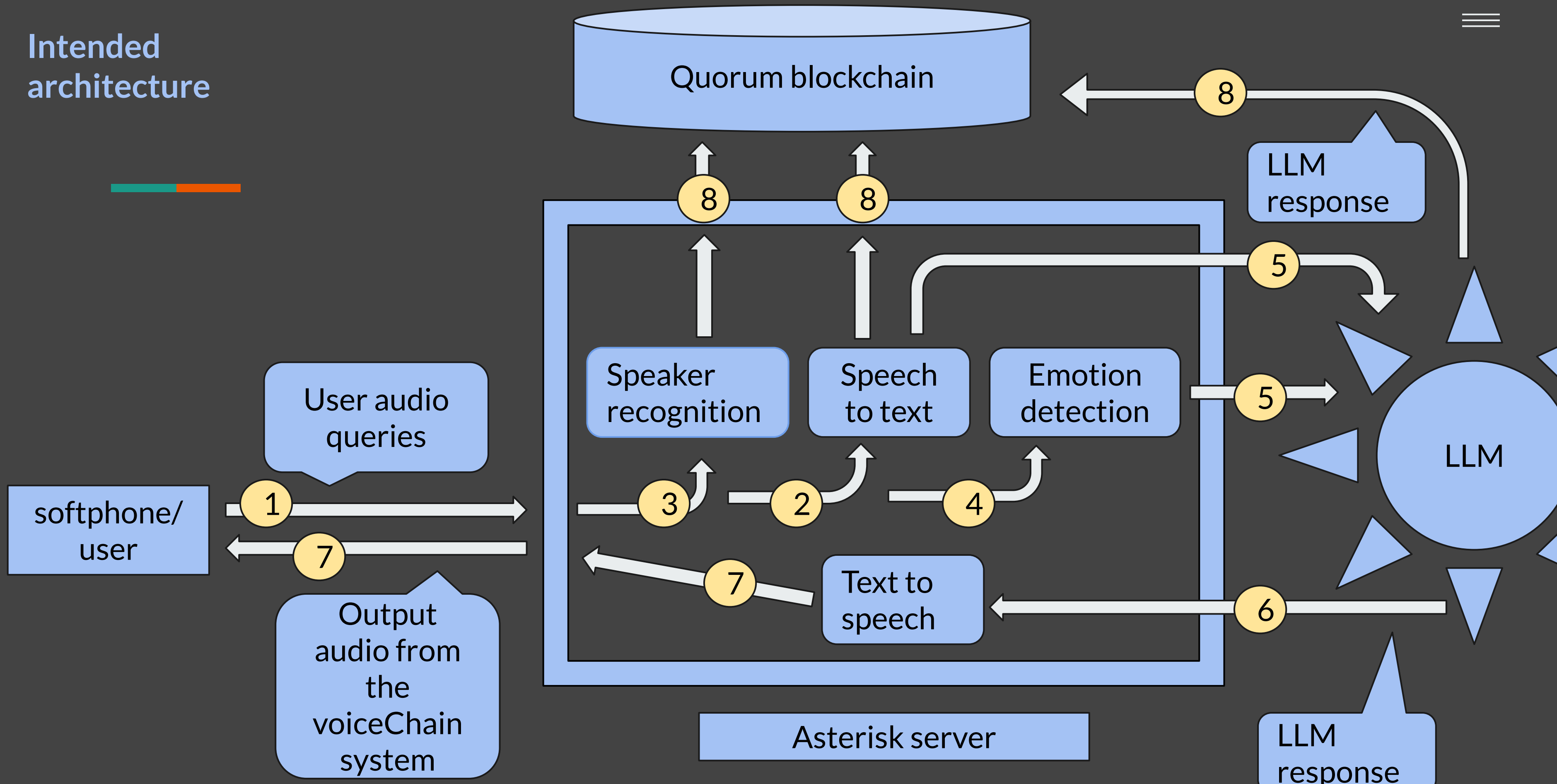
Emotion Detection: Analyses the emotional tone of the speaker's input for better response.

LLM Integration: Processes the converted text query using a Large Language Model to generate intelligent responses.

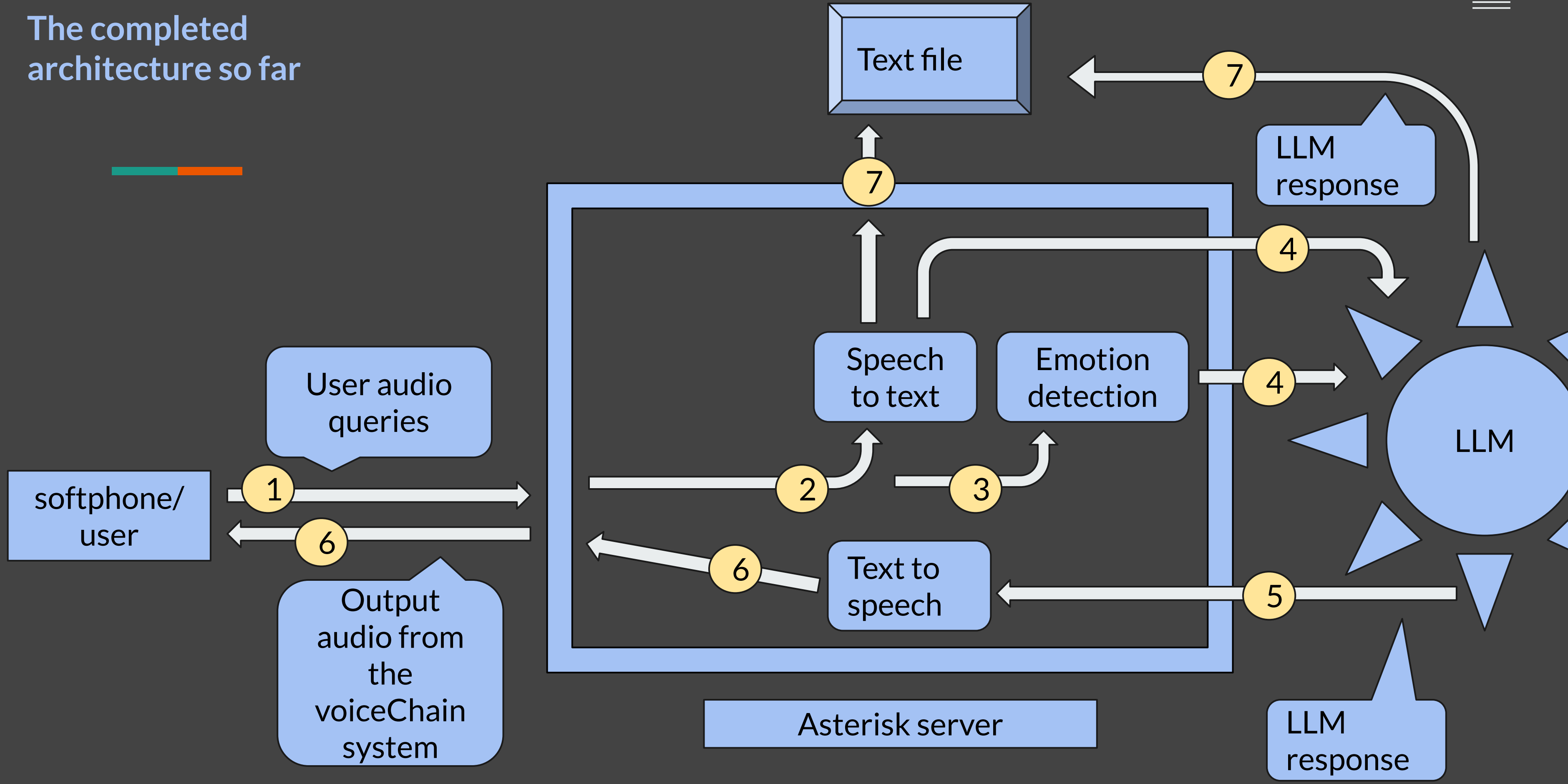
Real-time Feedback: Converts the LLM-generated response back into speech and provides feedback to the user.

System Architecture.

Intended architecture



The completed architecture so far



System Architecture Overview

Asterisk Server:

- Acts as the main server, handling incoming audio queries.
- Receives user input and manages communication between various components.

Speech-to-Text Conversion:

- Converts the incoming audio into a text format.
- This text is further processed to form the query sent to the LLM.

Emotion Detection:

- Emotion recognition from voice helps analyse the sentiment of the speaker’s input.
- Based on the emotional analysis, the LLM response can be adjusted.

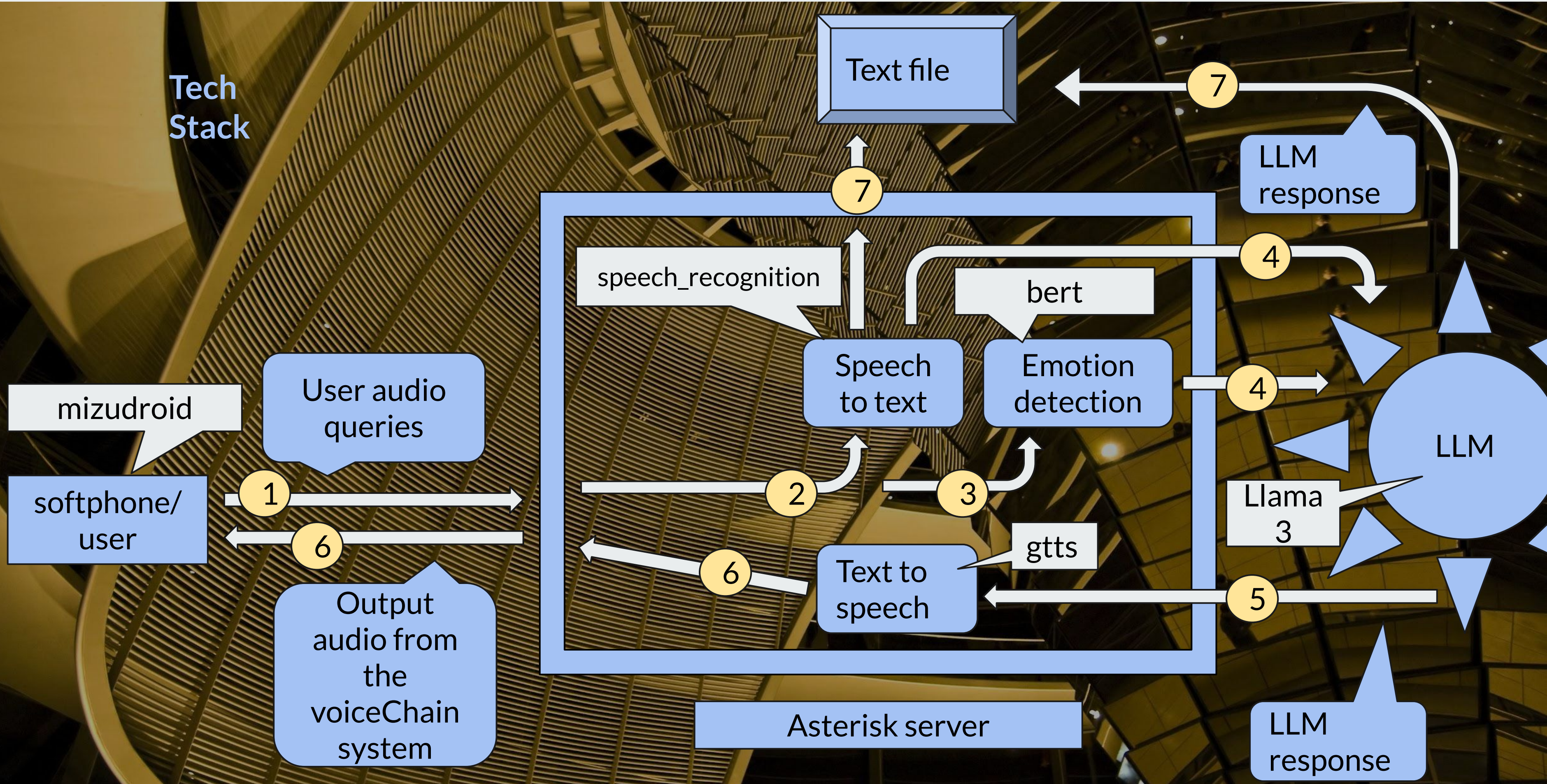
LLM (Large Language Model):

- processes the converted text query.
- provides an intelligent response based on the user's request and the knowledge provided to the LLM.

Text-to-Speech Conversion:

- The LLM’s text-based response is converted back into speech format for output.
- This allows for natural communication with the user.

Tech Stack


Text file
7
LLM response
7
speech_recognition
4
bert
Speech to text
Emotion detection
4
mizudroid
User audio queries
LLM
softphone/ user
1
2
3
Llama 3
6
6
gtts
Text to speech
5
Output audio from the voiceChain system
Asterisk server
LLM response

Tech stack overview

1. Asterisk

- A powerful open-source framework for building communication systems, Asterisk allows users to create VoIP services, including PBX (Private Branch Exchange), voice mail, and conferencing, making it ideal for handling telephony services.

2. LLaMA 3 (Large Language Model)

- LLaMA 3, developed by Meta, is a powerful language model built on transformer architecture. It excels at generating human-like text and can be fine-tuned for various NLP tasks such as text generation, summarization, and translation. LLaMA 3 is designed for efficiency and scalability, making it suitable for both research and production environments.

3. OOLaMA (Offline Optimized Language Model Assistant)

- OOLaMA is a platform designed for hosting and running large language models (LLMs) locally on a computer. It enables offline processing of LLM queries, providing privacy and security by eliminating the need for internet connectivity. Built for efficiency, OOLaMA supports various NLP tasks such as text generation, summarization, and question-answering, making it ideal for both research and private use cases.

3. SpeechRecognition

- A Python library for performing speech-to-text conversion, SpeechRecognition provides easy access to several popular speech APIs, including Google Web Speech, enabling users to convert spoken language into text efficiently.

4. GTTS (Google Text-to-Speech)

- GTTS is a Python library that provides an interface to Google Translate's text-to-speech API, gTTS can convert text into spoken language in multiple languages and is widely used in voice-enabled applications.

FOR Emotion detection

5. BERT (Bidirectional Encoder Representations from Transformers)

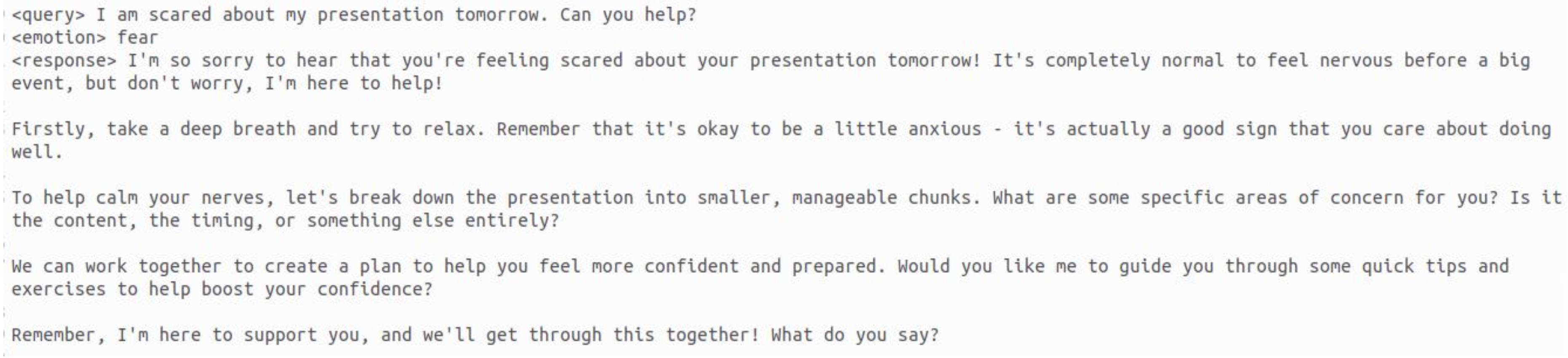
- A transformer-based model developed by Google, it will be used in VoiceChain for emotion detection, improving response accuracy by understanding the sentiment behind user queries.

6. Mizudroid

- Mizudroid is a SIP-based softphone application for Android devices. It allows users to make VoIP calls over the internet by connecting to a SIP server, such as Asterisk. In the **VoiceChain** system, Mizudroid serves as the client interface for users to interact with the IVR system.

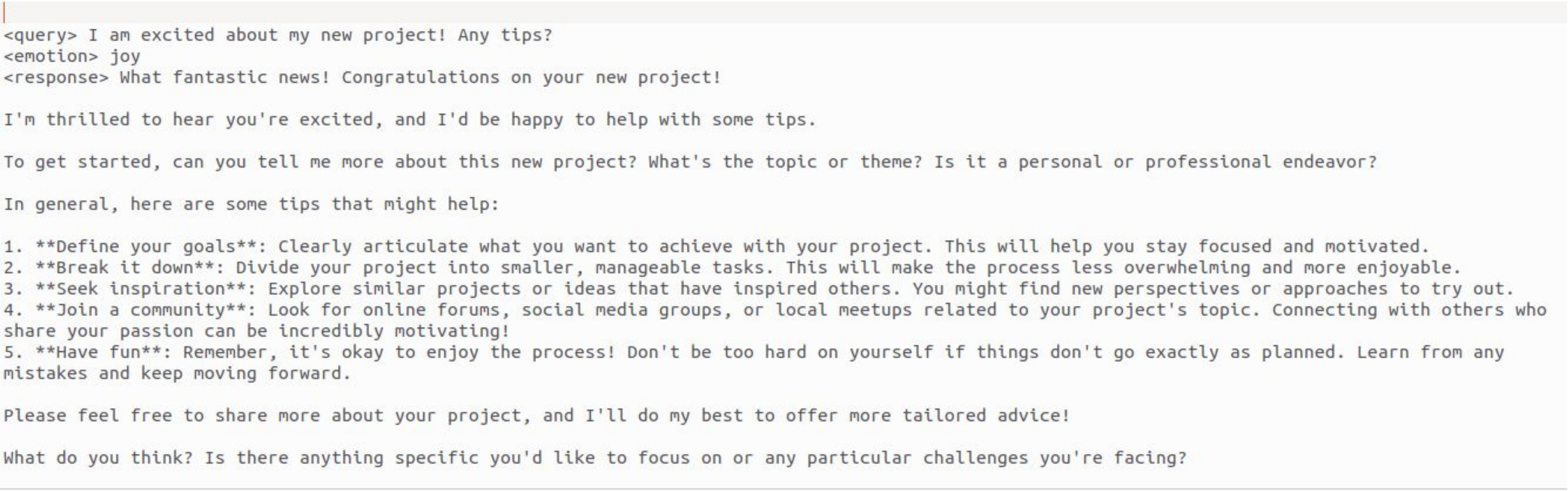
Experiments & Results

Sample responses from the VoiceChain system:



```
<query> I am scared about my presentation tomorrow. Can you help?
<emotion> fear
<response> I'm so sorry to hear that you're feeling scared about your presentation tomorrow! It's completely normal to feel nervous before a big
event, but don't worry, I'm here to help!

Firstly, take a deep breath and try to relax. Remember that it's okay to be a little anxious - it's actually a good sign that you care about doing
well.

To help calm your nerves, let's break down the presentation into smaller, manageable chunks. What are some specific areas of concern for you? Is it
the content, the timing, or something else entirely?

We can work together to create a plan to help you feel more confident and prepared. Would you like me to guide you through some quick tips and
exercises to help boost your confidence?

Remember, I'm here to support you, and we'll get through this together! What do you say?
```

More responses from System:



```
<query> I am excited about my new project! Any tips?
<emotion> joy
<response> What fantastic news! Congratulations on your new project!

I'm thrilled to hear you're excited, and I'd be happy to help with some tips.

To get started, can you tell me more about this new project? What's the topic or theme? Is it a personal or professional endeavor?

In general, here are some tips that might help:

1. **Define your goals**: Clearly articulate what you want to achieve with your project. This will help you stay focused and motivated.
2. **Break it down**: Divide your project into smaller, manageable tasks. This will make the process less overwhelming and more enjoyable.
3. **Seek inspiration**: Explore similar projects or ideas that have inspired others. You might find new perspectives or approaches to try out.
4. **Join a community**: Look for online forums, social media groups, or local meetups related to your project's topic. Connecting with others who
share your passion can be incredibly motivating!
5. **Have fun**: Remember, it's okay to enjoy the process! Don't be too hard on yourself if things don't go exactly as planned. Learn from any
mistakes and keep moving forward.

Please feel free to share more about your project, and I'll do my best to offer more tailored advice!

What do you think? Is there anything specific you'd like to focus on or any particular challenges you're facing?
```

Sample Audio Conversation with the VoiceChain system

. . .

Future work:

While the current system performs well in terms of speech recognition, emotion detection, and response quality, there are several avenues for further development:

- **Blockchain Storage:** The integration of blockchain for storing call records faces some challenges, particularly with Quorum's outdated resources and incomplete documentation. Future work will explore alternative blockchain platforms or address these issues to ensure secure and scalable data storage.

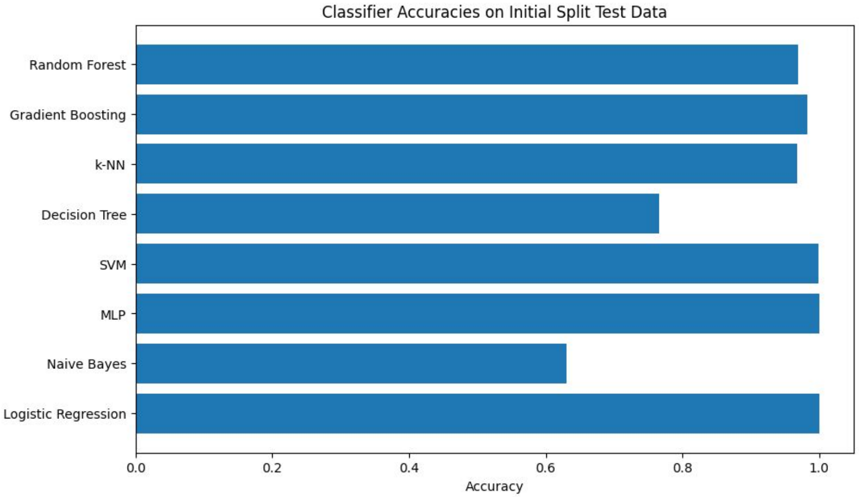
- **Comparative Emotion Analysis with DWFormer:** As part of ongoing improvements, the system will undergo a comparative emotion analysis with DWFormer, a model known for its capability of finding emotion from audio. This will be beneficial for further improving the accuracy of emotion recognition and better response in challenging environments.

- **Speaker Recognition:** Future work will focus on implementing speaker recognition capabilities to personalize interactions further. This will allow the system to recognize and respond differently based on the individual user, improving the user experience.

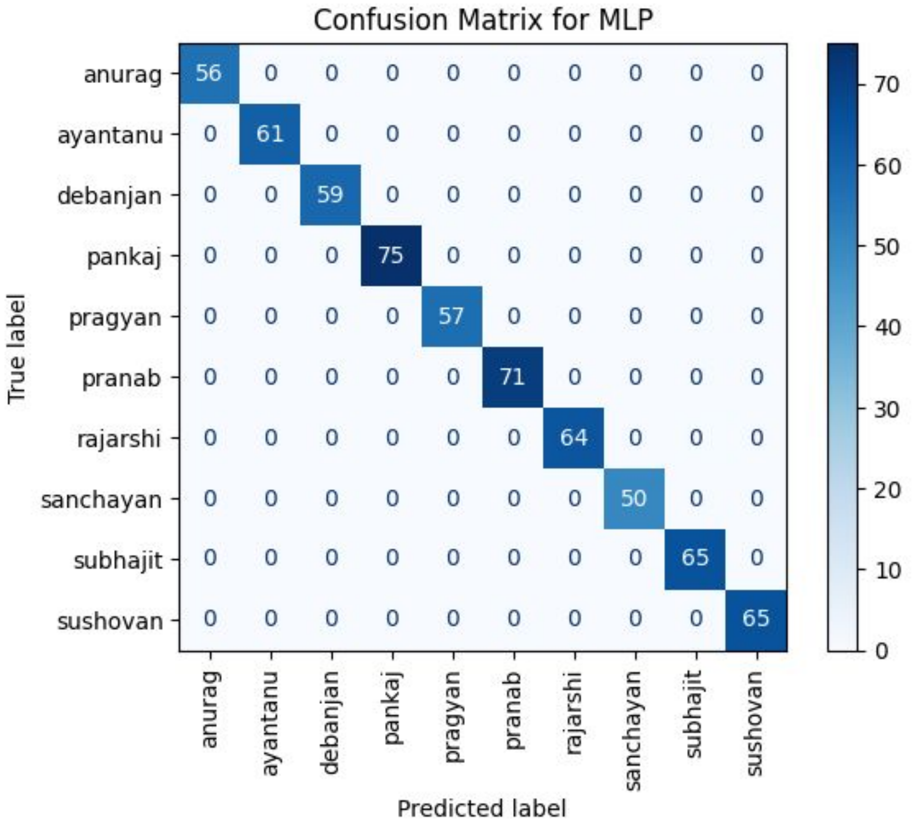
Output of speaker recognition code on custom dataset. (unfinished)

There are 10 individual speakers, with each having close to 300 audio recordings(1 second each). In my code, I had performed an 80-20 split for the train-test split, then used various machine learning algorithms.

```
nitial Split Test Accuracies:
Random Forest: 96.9%
Gradient Boosting: 98.2%
k-NN: 96.7%
Decision Tree: 76.5%
SVM: 99.8%
MLP: 100%
Naive Bayes: 63.08%
Logistic Regression: 100%
```



Confusion matrix of the best classifier



Confusion Matrix for MLP

True label \ Predicted label	anurag	ayantanu	debanjan	pankaj	pragyan	pranab	rajarshi	sanchayan	subhajit	sushovan
anurag	56	0	0	0	0	0	0	0	0	0
ayantanu	0	61	0	0	0	0	0	0	0	0
debanjan	0	0	59	0	0	0	0	0	0	0
pankaj	0	0	0	75	0	0	0	0	0	0
pragyan	0	0	0	0	57	0	0	0	0	0
pranab	0	0	0	0	0	71	0	0	0	0
rajarshi	0	0	0	0	0	0	64	0	0	0
sanchayan	0	0	0	0	0	0	0	50	0	0
subhajit	0	0	0	0	0	0	0	0	65	0
sushovan	0	0	0	0	0	0	0	0	0	65

This speaker recognition model behaves not well enough in the VoiceChain system due to audio compression over call and noise over call, so it is not in the current system; it needs further improvements.

Use cases

- **Customer Support**: Automated IVR systems using LLMs can intelligently handle customer queries and route them to the correct department or provide immediate responses.
- **Call Centers**: The secure handling of sensitive data in call records is crucial for maintaining compliance with regulations like GDPR.
- **Telemedicine**: A secure platform that can recognize the speaker, process medical-related queries, and store important call records in a tamper-proof environment.
- **Legal and financial consultation**: ensures secure and personalized communication while securely storing all interactions on a blockchain for future reference.

Benefits of VoiceChain

Efficiency: LLM-powered responses reduce the need for human intervention in customer service processes.

Real-time Interaction: Natural language processing through speech-to-text and text-to-speech systems ensures a seamless conversation between the system and users.

Scalability: The system is designed to handle a large number of users, making it ideal for organisations with high call volumes.

Data Privacy: All personal data and call records are securely stored, ensuring compliance with privacy regulations.

conclusion

VoiceChain offers a highly scalable, secure, and intelligent IVR solution for modern businesses. By integrating emotion detection, LLMs, ensure that all interactions are handled efficiently. This system is ideal for industries that require secure, intelligent handling of customer queries, and it provides a robust solution to modern IVR challenges.

Thank you.