

**Building a Virtual Assistant Using Deep Learning and LLM to Enhance the Trading Experience
for Seniors & People with Disabilities**

Project 3 Milestone 3

Guruprasad Velikadu Krishnamoorthy

College of Science and Technology, Bellevue University

DSC680-T301: Applied Data Science

Professor. Amirfarrokh Iranitalab

August 10, 2024

Building a Virtual Assistant Using Deep Learning and LLM to Enhance the Trading Experience for Seniors & People with Disabilities

1. Project Description:

The Project aims to enhance the Stock Trading experience by building a Virtual Assistant using LLM (Large Language Model) and Langchain technologies to make Mobile Stock Trading more accessible for Seniors and People with specific Disabilities, such as the Visual and hearing impaired. It also enhances the application's security by implementing Bio-Metric Authentication using Voice Recognition and Lip-Reading using Neural Networks and Deep Learning. Overall, the project intends to enhance the experience by limiting or eliminating the need for typing while accessing the trading platform, thus not requiring entering or remembering the credentials.

2. Background:

With the rise of Technology and everything becoming digital, most sectors, such as Restaurants, Utilities, Health care, Banking, and even the Government, want us to use Apps on our smartphones. Using mobile apps on the phone can be very convenient, especially for banking and utilities. It can provide many benefits, such as making payments on the go, skipping the long wait to talk to a representative, etc. The rise of commission-free stock trading apps like Robinhood, M1 Finance, etc., has completely redefined the trading experience, allowing us to buy and sell stocks in the comfort of our homes without paying a penny in brokerage fees.

All that was well until the corporations and app makers started to forget and neglect a part of society that was still struggling to do basic things like accessing the internet and using

smartphones. According to a study by Pew Research (Faverio, 2022), only 61% of Americans aged 65 and above owned a smartphone as of 2021, and the number was only 46% back in 2018. The pandemic has accelerated the trend and even forced some seniors to use smartphones for the first time.

The other section of society that is mostly forgotten is the people with some disability. According to a study from Deloitte, about 67 million Americans have some disability, which represents one-fifth of the U.S. population. Also, the study found that most disabled people felt the banking or Trading experiences were not accessible at all. People with extreme disabilities, such as the visually impaired and hearing impaired, are the most affected and can find the traditional mobile banking experience cumbersome. (*How—and Why—Banks Can Better Serve People with Disabilities*, n.d.)

3. Business Problem:

- I. **Too much Technology!** The research found that most people ages 65 and above surprisingly enjoyed the Digital experience, though they were cautious about using Banking and Trading apps. This is mainly because they were worried about tapping the wrong button, getting locked out, clicking an incorrect option, and losing all their money. Most mobile trading and banking apps provide the same interface for all age groups, and many are crowded with too many choices, which can easily confuse a first-time and not-so-digital-savvy user.
- II. **Remembering the password is a struggle!** Compared to young people, another problem that old adults face is password management. With almost every app on the phone requiring creating a user ID and password, maintaining too many credentials and

remembering them can take time and effort. Though there are various options, such as digital wallets, to store and generate a unique password, this will add another layer of complexity to the mix. Keeping the same username and password can be easily exploited if it gets into the wrong hands. Face ID biometric authentication is another option; however, many older people either don't trust them or have difficulty setting it up on their phones. Hence, it is up to banks and brokerage platforms to provide a secure solution and boost older adults' confidence that their money is safe and that they can trade and bank easily and securely without fear of losing it all. (McEachran, 2021)

III. People with Disabilities are also struggling! People with hearing and Vision disabilities face a similar struggle. Though the screen reading options can help them navigate the options on the screen and the Braille keypad can help type on the screen, the whole experience can be very challenging. Security can be another challenge. Face ID Biometric authentication may not be most effective for Blind and Visually impaired customers.

Overall, the problems faced by both forgotten sections of society- the older population and people with disabilities are the same. The need for easy navigation using mobile apps and easy and secure biometric authentication prevents or limits them from using mobile trading and banking apps on smartphones and tablets. ***This is the Business problem being addressed in this Project.***

There is so much untapped Potential! According to bankrate.com, the average account balance in the bank of people ages 65 and above is about 100k. Most funds either stay in the savings or checking account or prefer to invest in traditional options such as Bonds or high-interest CDs, etc. Any money invested in the stock market is usually made with the help of a portfolio manager, who

manages the funds for a commission. *Hence, this forgotten part of society has so much untapped potential. (Bennett, 2024)*

4. How can Machine Learning help?

- **Authentication:** Setting up Face ID authentication can be challenging for some. Hence, in this Project, biometric authentication is implemented by using the customer's voice as the password. When the customer calls the bank or opens the mobile app, they will be asked to repeat the numbers from 0 through 9. By analyzing the voice commands using deep learning and Neural networks, the Model can identify if the real customer is trying to access them using Voice Identification. The customer will also be asked to enter a Passphrase to provide another layer of Protection. They can either type it or face the camera and silently whisper or read out the passphrase of their choice. Lip reading using deep learning and neural networks will validate the passphrase and confirm whether the person is an actual user. It is worth noting that Lipreading is recommended instead of capturing the audio and converting it to Text, thus keeping the passphrase secure and avoiding exploitation by someone overhearing or while accessing the application in a loud background or in public.
- **Virtual Assistant:** Once the customer is authenticated, to make the trading experience more accessible, they can interact with the virtual assistant, asking simple questions in English, such as "What is my current Balance?" or "What is the price of Apple Stock today?". They will either see the response on the screen or be prompted if they want to hear the response. Data is stored in the MySQL Database, and the calls will be made using large language models such as Google Palm in combination with Lang Chain to accomplish this.

The speech will be converted to Text using the models provided by the Whisper library in Python. Also, a Streamlit app was built to display the demonstration results.

To Summarize, the six modules that were built in this Project are:

- I. Speaker Identification or Voice Recognition using the LSTM Model and Deep Learning to provide Biometric Authentication
- II. Lip Reading uses CNN and LSTM to provide a second layer of biometric authentication to capture the Passphrase.
- III. Speech to Text conversion using Whisper Models if the customer chooses to speak to the Virtual Assistant
- IV. Using Google Palm and Langchain to build a virtual assistant that provides an interactive response to the questions asked using the natural English language.
- V. A MySQL database containing sample data to which the calls will be made.
- VI. Streamlit app to show a simple demonstration of the Virtual Assistant.

The Flow diagram of the different modules is shown in the figure below:

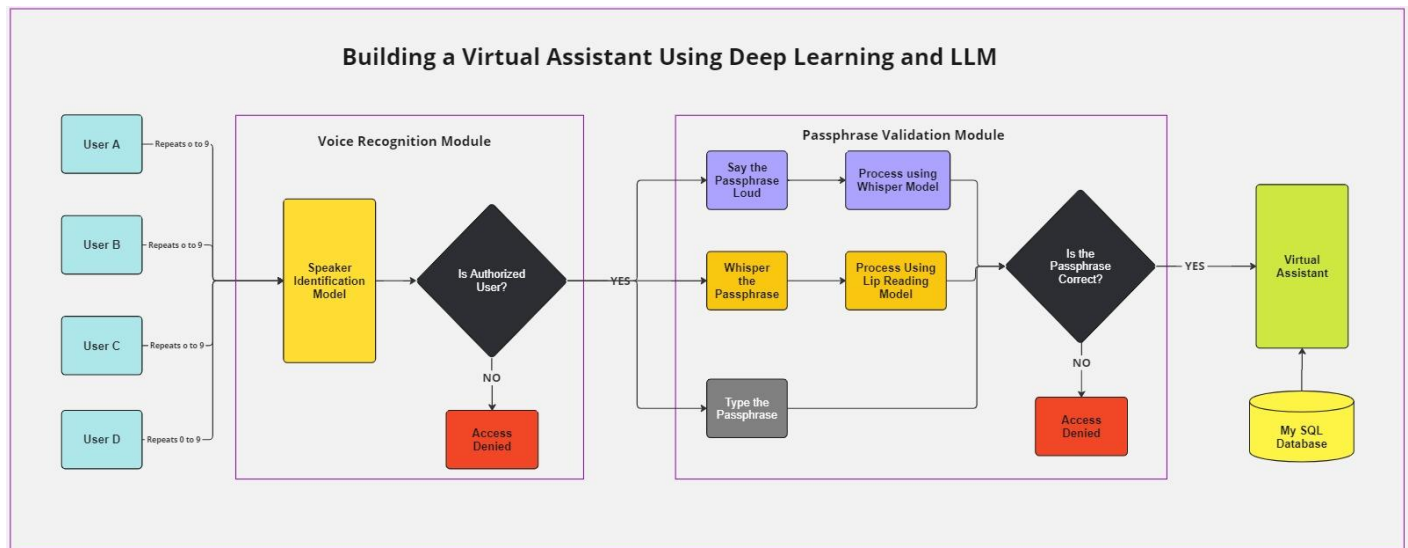


Figure1: Flow diagram of the different modules

5. Datasets:

The dataset used for lipreading was taken from the GRID audiovisual sentence corpus, which contains a vast collection of audio and videos to support behavioral studies on speaker perceptions. Though the entire corpus contains the videos recorded by 34 actors, each stored in a separate folder, in this Project, the videos of one actor were used. Every folder includes 1000 videos, each lasting a few seconds long. Also, an alignment file consists of the Text spoken in the video. (*The GRID Audiovisual Sentence Corpus*, n.d.)

The Free-Spoken Digit Dataset for Speaker Recognition was taken from GitHub and contains 3000 recordings of the digits spoken by six actors. (*Free-Spoken-Digit-Dataset/Recordings at Master · Jakobovski/Free-Spoken-Digit-Dataset*, n.d.)

To create a mock stock portfolio mimicking a real-world use case, dummy data was produced and loaded into the MySQL Database.

6. Methods and Analysis:

6.1. Voice Recognition using Speaker Classification Model – Module 1

When older customers call customer support or log into the app, their voice will be used as the password by analyzing the patterns in the voice, thus providing biometric authentication, which does not require remembering or entering the user ID or password. As multiple authorized Users can be in an account, the model must identify more than one voice per account. To accomplish this, a model was built by analyzing the voices of multiple users, and only two among the many speakers were considered authorized users. The customer would be asked to read out numbers from zero through Nine, which will be analyzed, and access will be granted. The model will deny entry if anyone other than the authorized user tries to access the account.

6.1.1. Data Exploration:

The dataset for speaker classification contained the voice recordings of six actors speaking the digits from zero through nine, fifty times each. There were 500 recordings for each speaker, as shown in the figure below.

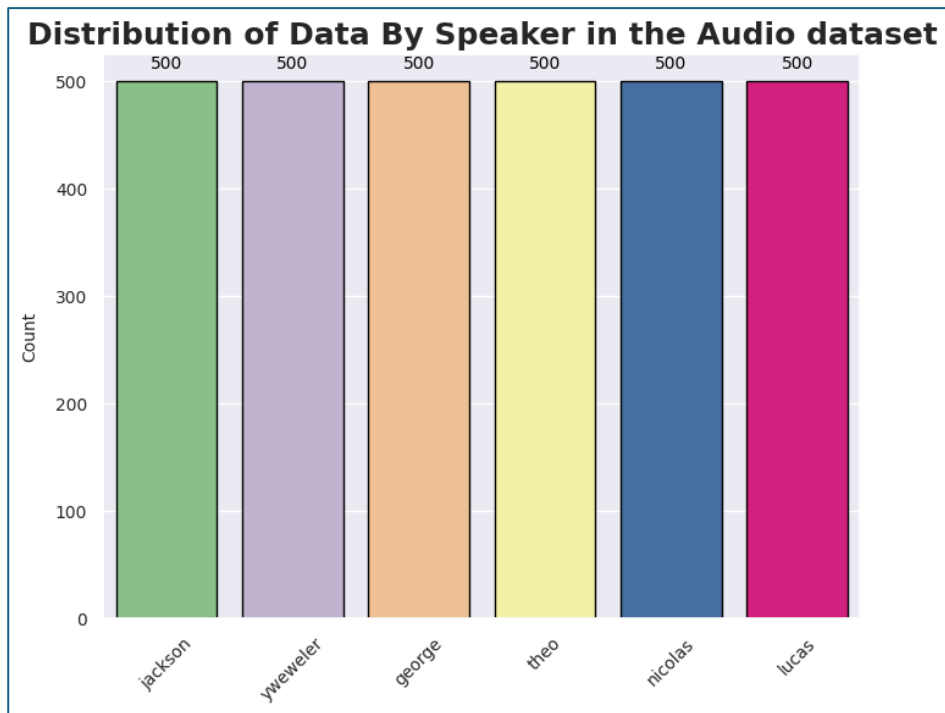


Figure 2: Distribution of data in the Audio Dataset

The audio files in the .wav format were converted into Numeric form to build the Machine learning models using Librosa Python Library. The figure below displays a sample Wave plot showing the distribution of the audio signals.

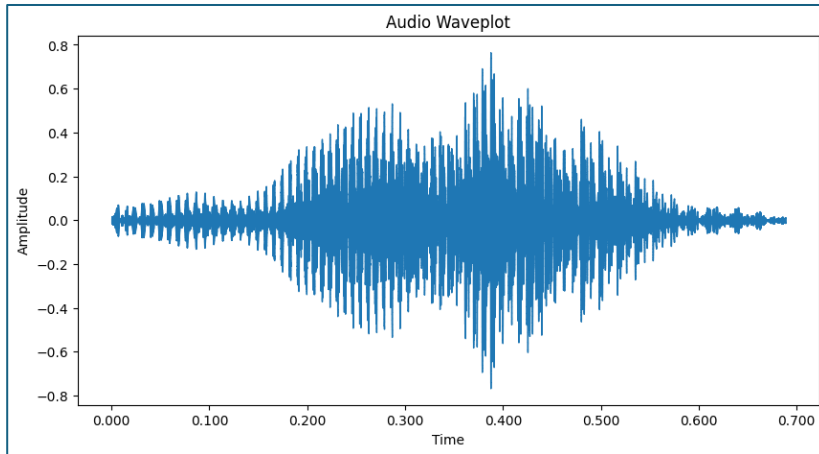


Figure 3: Wave Plot of a sample Audio file

6.1.2. Data Preparation:

As the model requires the users to read the numbers from zero through Nine, the individual audio files were combined based on the speaker and the Iteration count. As there were 50 iterations for each speaker, the entire dataset was scanned to form the combined dataset for each speaker, thus creating 300 combined audios from 3000 individual files.

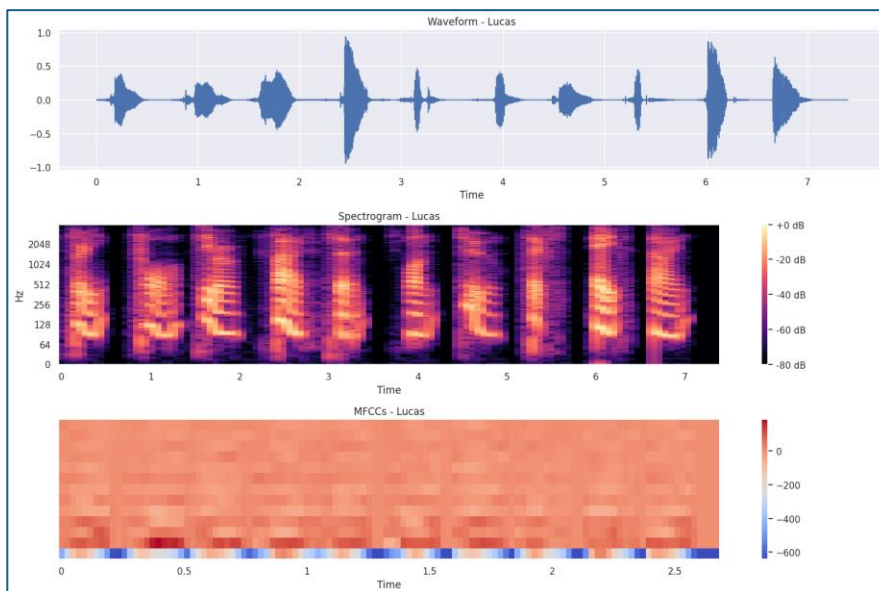


Figure 4: Data Exploration of a combined file for Speaker Lucas

For instance, the file **2_george_0.wav** represents the second iteration file for the user "George," speaking the numbers zero to Nine. Ten such audio files of George speaking the numbers from the second iteration were concatenated to form a combined file. Figures 4 and 5 include Wave Plot, Spectrogram, and MFCC representation of the combined audio files for Lucas and Jackson speakers.

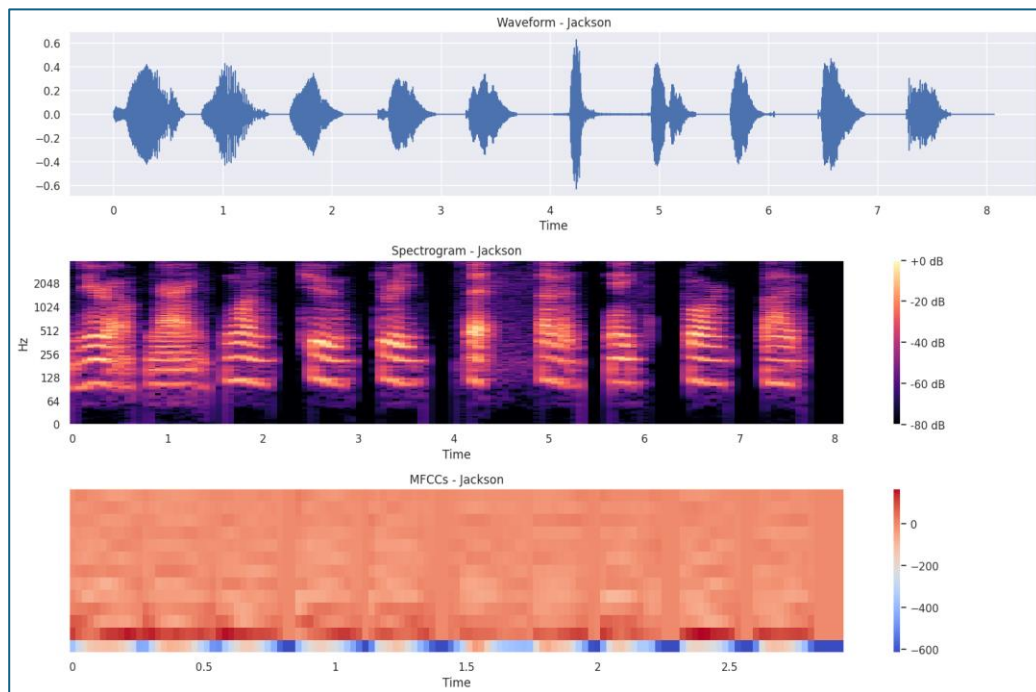


Figure 5: Data Exploration of a combined file for Speaker Jackson

6.1.3. Feature Extraction and Model Building:

The MFCC (Mel- Frequency Cepstral Coefficient), which represents the audio signals in the Frequency domain, was extracted from the combined audio files. The MFCC plot shown in the figures above includes Time versus different Frequency bands, with each column representing short segments of time and the height representing the intensity or magnitude of the signal in the band.

Feature extraction is a critical step in the data preparation phase of machine learning projects as it can help with various aspects such as dimensionality reduction, noise reduction, and Model performance improvement. As not all the raw data can be equally valuable, the feature extraction helps only retain the relevant aspects of the data. 13 MFCC coefficients were extracted from the audio data, with each coefficient representing data such as spectral slope, shape of vocal tract, nasality of speech, etc.

The extracted features in a numeric array were standardized using a standard scaler and then padded to maintain a fixed length. This process was repeated for all 300 combined files created in the data preparation step. The target classes comprising six speakers were labeled encoded, and the dataset was split into Train, Test, and Validation sets.

The LSTM (Long Short-Term Memory) model, a Recurring Neural Network (RNN) model that works well with Sequential data, was used in this project. LSTMs can retain the information for much longer than traditional RNN methods, so it was a better choice for speaker and voice classification algorithms. (Sugandhi, 2023).

The LSTM Layer with 128 neurons was used, while the “ReLU” (Rectified Linear Unit) was used as an activation function in the hidden layer with 64 neurons. The softmax function was used as an activation function ideal for multi-classification models in the final output or dense layer with as many units as the number of speakers. Also, early stopping was used to avoid overfitting, and the model was trained on 20 epochs using an "Adam" optimizer for compilation. The Sparse Categorical Cross Entropy was used as a Loss function, which is used when the target class is represented as an integer class instead of a one-hot encoded variable.

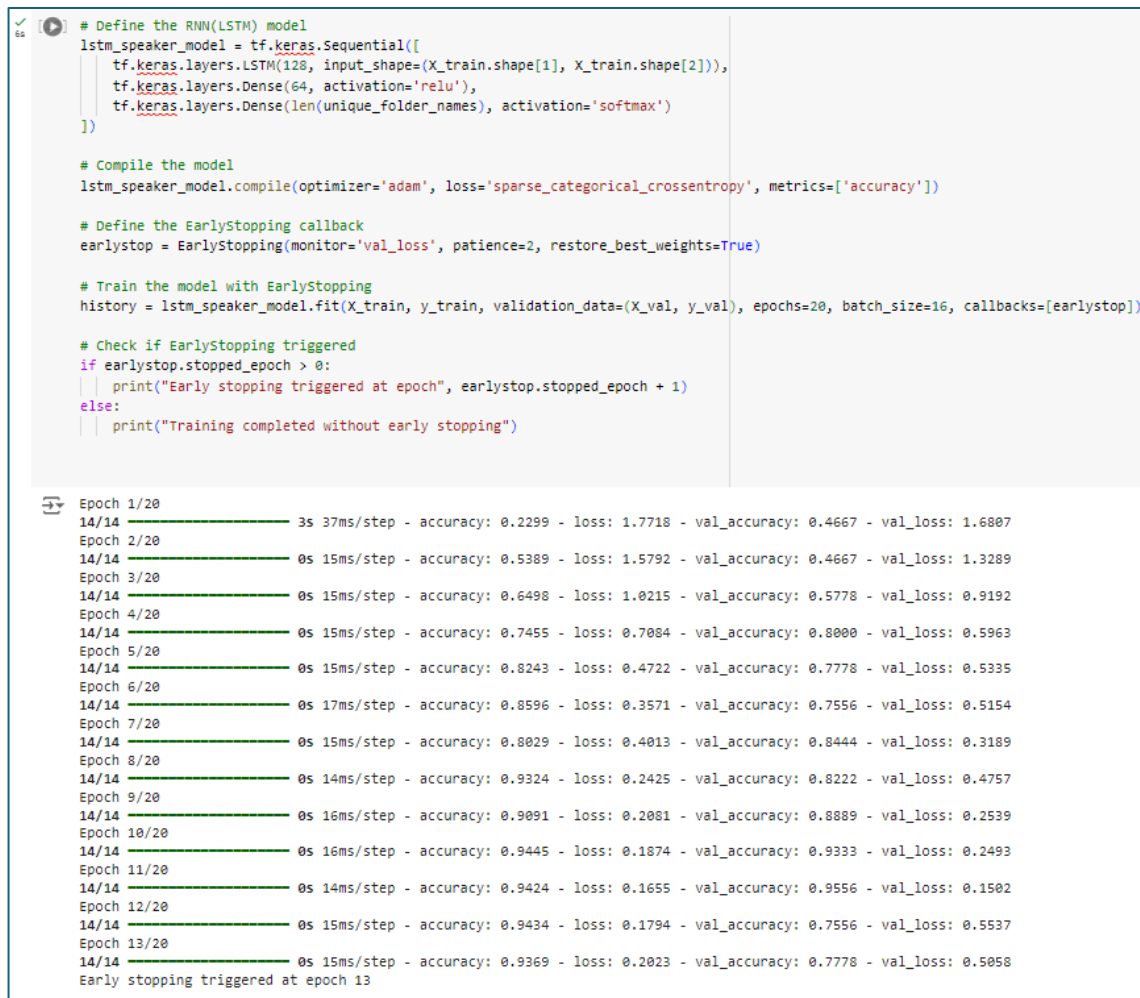


Figure 6: LSTM Model for Voice Recognition

6.1.4. Evaluation of Voice Recognition Model:

The Voice Recognition model was evaluated by plotting the Loss, Accuracy, and Confusion Matrix. Accuracy represents the percentage of predictions where the predicted value is the same as the actual value. From the plot, the Accuracy appears to increase as the number of Epochs increases, while the Loss, which represents the sum of errors made in each sample, decreases as the number of epochs increases.

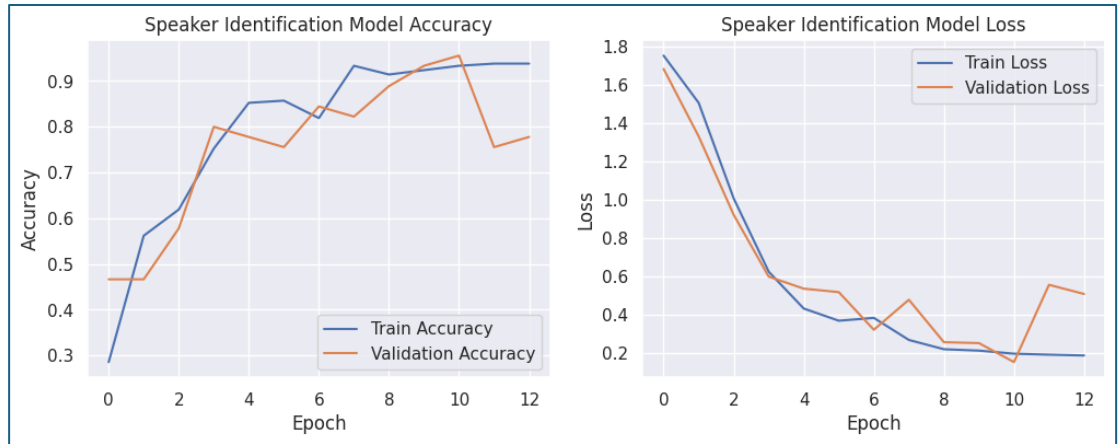


Figure 7: Loss and Accuracy Plots for Voice Recognition Model

The Confusion Matrix, commonly used to describe the Classification model performance, suggests that the Model predicted most values correctly. The test evaluation accuracy of the model was 86%, indicating the model could predict most values correctly.

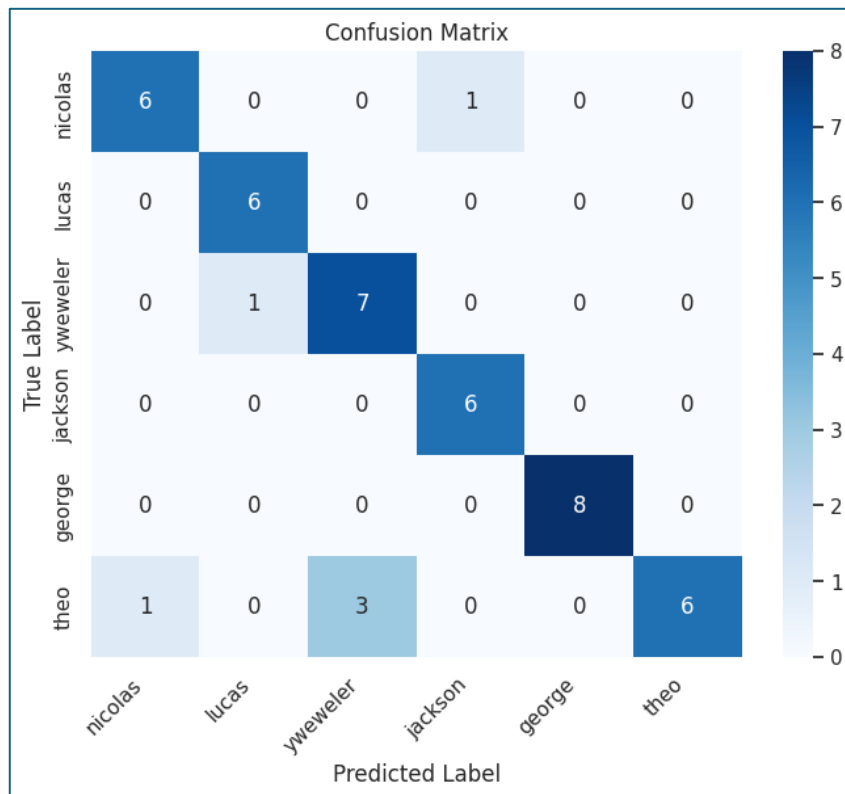


Figure 8: Confusion Matrix of Voice Recognition Model

Passphrase identification

To provide a second layer of Protection, the customer would be prompted to enter or say the passphrase to access the account. The customers would be given multiple choices, such as typing the passphrase or saying it if they are in a safer environment or not in public. The Whisper model was used to convert the audio to Text and validate the passphrase. If the customer feels uncomfortable saying it out loud, they can also choose to whisper the passphrase. Lip reading was used to analyze the lip movements to identify the passphrase.

6.2. Whisper Model- Module 2

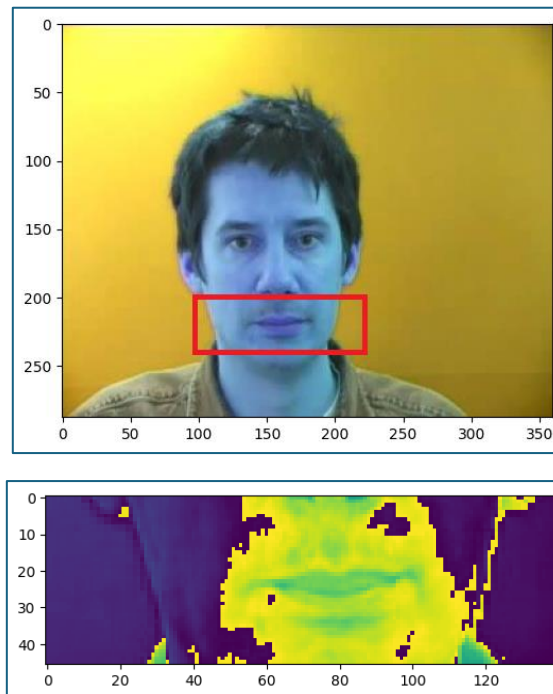
The Whisper is an open-source Speech recognition model built by OpenAI that can convert audio signals from various languages into English. In this project, the whisper model was used to convert the passphrase spoken by the customer into Text, and the results were then validated and authenticated. (*Whisper (Speech Recognition System)*, 2024).

6.3. Lip Reading Model using Deep Learning- Module 3

To train and build a Deep learning model to identify the passphrase, the dataset was downloaded from the GRID audiovisual sentence corpus, which contains a vast collection of Audio and Videos to support Behavioral studies in speaker perceptions. Though the entire corpus contains videos recorded by 34 actors, one actor was used in this project.

6.3.1. Data Exploration and Preparation:

The video files did not require any cleansing as the video files and alignment files containing the Text spoken were labeled clearly. Multiple Image frames from the Video files were extracted using the cv2 library, and the area around the lips was extracted for further processing. A sample image file is shown below, highlighting the image section used for processing.



*Figure 9: Sample image extracted from a video file
highlighting the section of the mouth*

As the dataset contains multiple videos and alignment files, they were extracted simultaneously for each combination by parsing the entire dataset. The video files were then scanned frame by frame and converted into Numpy arrays. The numpy arrays for all frames were concatenated to form a combined array for each video file. Similarly, the text of the alignments file was extracted and converted into numbers.

```
[ ] def create_frames_alignments(video_file_path):
    """
    Function to convert the video files by parsing it frame by frame and
    creating a collection of Numpy array. Only the area around the
    Person's mouth is used for further processing.
    The extracted text from the alignments file will be converted to numbers
    """
    video_file_path = bytes.decode(video_file_path.numpy())
    extracted_annotation=video_file_path.split("/")[-1].split(".")[0]
    video_file_path=os.path.join("/content/s1",extracted_annotation + ".mpg")
    alignment_file_path=os.path.join("/content/align",extracted_annotation + ".align")
    std_video_frames=convert_video_to_frames(video_file_path)
    num_alignments=extract_alignments(alignment_file_path)
    return std_video_frames,num_alignments

# Calling the function to convert the video and alignments text into Numbers
std_video_frames,num_alignments = create_frames_alignments(tf.convert_to_tensor(sample_video_path))
```

Figure 10: Screenshot of the function used to convert Video files and Alignments into Array

A sample of 100 video files was used for training the models, and each video file, along with its alignment file, was converted into a NumPy array. The dataset was split into Train and Test sets in the 80:20 ratio and converted into a tensor dataset before model building.

6.3.2. Model Building and Evaluation:

A Convolutional 3D Model with multiple hidden layers, each with 128, 256, and 75 neurons, was used, and ReLU was used as an activation function in these layers. The data was then flattened, and two more layers of Bidirectional LSTM were added to the model, as shown in the figure. Unlike traditional RNNs that can read in one direction, the bidirectional LSTM captured the context in both directions. The final dense layer uses Softmax as an activation function and the maximum vocabulary size (40 in this model) as the number of classes. The model was then trained on the Training dataset, and the results were validated using the Test data.


```

# Creating a model calling the Sequential API
lstm_model = Sequential()
# Adding 128 neurons for the first Layer
lstm_model.add(Conv3D(128, 3, input_shape=(75,46, 140,1), padding='same'))
# Using Rectified Linear Unit as Activation
lstm_model.add(Activation('relu'))
lstm_model.add(MaxPool3D((1,2,2)))

# Adding second layer with 256 Neurons
lstm_model.add(Conv3D(256, 3, padding='same'))
lstm_model.add(Activation('relu'))
lstm_model.add(MaxPool3D((1,2,2)))

# Adding third layer with 75 neurons matching the number of Frames
lstm_model.add(Conv3D(75, 3, padding='same'))
lstm_model.add(Activation('relu'))
lstm_model.add(MaxPool3D((1,2,2)))

# Flattening the data
lstm_model.add(TimeDistributed(Flatten()))

# Adding a Bidirectional LSTM Layer with dropout
lstm_model.add(Bidirectional(LSTM(128, kernel_initializer='Orthogonal', return_sequences=True)))
lstm_model.add(Dropout(.5))

lstm_model.add(Bidirectional(LSTM(128, kernel_initializer='Orthogonal', return_sequences=True)))
lstm_model.add(Dropout(.5))

#Final Dense Layer with the Softmax as the activation layer
lstm_model.add(Dense(alphanumchar_to_number.vocabulary_size(), activation='softmax'))

```

Figure 11: Screenshot of the CNN and LSTM models used for Lip Reading

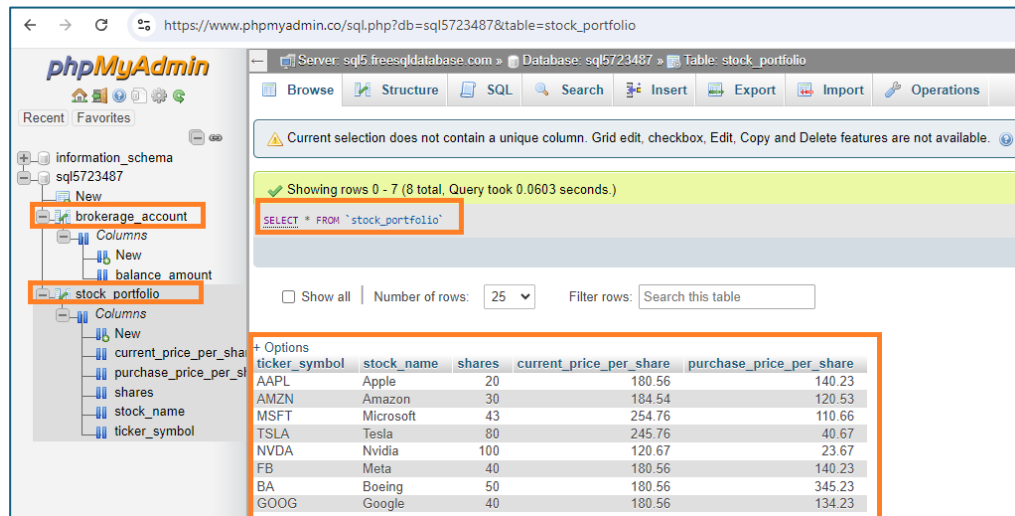
Building a Virtual Assistant using LLM and Langchain

After the customer is authenticated using voice recognition and Passphrase validation, they can access the Trading account and perform further transactions. To make the trading experience more accessible, a virtual assistant was built so that the customer could ask questions about the Portfolio using simple English words. The model would return the results by fetching the data from the backend database.

6.4. Creating Database and Tables in MySQL- Module 4

To mimic the real-world scenario of a customer's stock portfolio, two tables were created in this project, one containing the details about their stock portfolio and the other containing the balance amount. These were created in the MySQL Databases hosted on FreeSQLDatabase.com. The figure

below shows the query executed on the stock portfolio table containing the number of shares and their current and Purchase price details.



The screenshot shows the phpMyAdmin interface for a database named 'sql5723487'. The 'stock_portfolio' table is selected. The SQL query entered is 'SELECT * FROM `stock_portfolio`'. The results show 8 rows of data. The table structure includes columns: ticker_symbol, stock_name, shares, current_price_per_share, and purchase_price_per_share.

ticker_symbol	stock_name	shares	current_price_per_share	purchase_price_per_share
AAPL	Apple	20	180.56	140.23
AMZN	Amazon	30	184.54	120.53
MSFT	Microsoft	43	254.76	110.66
TSLA	Tesla	80	245.76	40.67
NVDA	Nvidia	100	120.67	23.67
FB	Meta	40	180.56	140.23
BA	Boeing	50	180.56	345.23
GOOG	Google	40	180.56	134.23

Figure 12: Screenshot of sample SQL Query results from the Database

6.5. Building the Virtual Assistant- Module 5

The Virtual Assistant was built using GooglePalm, a Large Language Model from Google that uses Transformers and can perform various reasoning tasks such as common reasoning, explaining the code, etc. (Wikipedia Contributors, 2024). In this project, Google Palm was integrated to work with the data from the online database using SQL Database Chain. When the user asks questions about the Portfolio, such as "What is my Balance amount?" the model can convert the user question into SQL queries, execute them in the backend database, and return the results.

The figure below shows a few sample questions asked by the customer and the results of the SQL queries.

```
> Entering new SQLiteDatabaseChain chain...
What is my current balance amount?
SQLQuery:SELECT balance_amount FROM brokerage_account
SQLResult: [[1000.0,]]
Answer:1000.0
> Finished chain.

> Entering new SQLiteDatabaseChain chain...
How many shares of Apple stock are in my portfolio?
SQLQuery:SELECT shares FROM stock_portfolio WHERE stock_name = 'Apple'
SQLResult: [[20,]]
Answer:20
> Finished chain.

> Entering new SQLiteDatabaseChain chain...
The what is the current price of Nvidia stock?
SQLQuery:SELECT current_price_per_share FROM stock_portfolio WHERE ticker_symbol = 'NVDA'
SQLResult: [[120.67,]]
Answer:120.67
> Finished chain.

> Entering new SQLiteDatabaseChain chain...
What is the best performing stock in my portfolio?
SQLQuery:SELECT stock_name FROM stock_portfolio ORDER BY (current_price_per_share - purchase_price_per_share) DESC LIMIT 1
SQLResult: [[('Tesla',)]
Answer:Tesla
> Finished chain.

> Entering new SQLiteDatabaseChain chain...
Do I have sufficient balance in my brokerage account to purchase five shares of Apple's stock?
SQLQuery:SELECT balance_amount FROM brokerage_account WHERE balance_amount >= 5 * (SELECT current_price_per_share FROM stock_portfolio WHERE ticker_symbol = 'AAPL')
SQLResult: [[1000.0,]]
Answer:Yes
> Finished chain.
```

Figure 13: Sample Responses for the Questions asked by the customer

6.5.1. Integration of Virtual Assistant with Whisper Model:

To provide capabilities for the customer to interact with the Virtual Assistant by asking questions instead of typing them manually, the model was also integrated with the Whisper Model. The Audio data will be converted into Text using the model, which will then be fed to the SQL Database chain, as explained in the above section. The figure below shows a few sample audio files containing the questions asked by the user that were converted into Text using the Whisper model.

```
✓ 2s [8] # Importing the Whisper Library
import whisper
whisper_model = whisper.load_model("base")

✓ 6s [9] # Extracting the text from the Audio File using Whisper Library
enquiry1_path='/content/Question1.wav'
whisper_result1=whisper_model.transcribe(enquiry1_path, fp16=False)
portfolio_qn1=whisper_result1['text']
print(f" The Extracted text from the Audio: {portfolio_qn1}")

→ The Extracted text from the Audio: What is my current balance amount?

✓ 6s [10] # Extracting the text from the Audio File using Whisper Library
enquiry2_path='/content/Question2.wav'
whisper_result2=whisper_model.transcribe(enquiry2_path, fp16=False)
portfolio_qn2=whisper_result2['text']
print(f" The Extracted text from the Audio: {portfolio_qn2}")

→ The Extracted text from the Audio: How many shares of Apple stock are in my portfolio?

✓ 10s [11] enquiry3_path='/content/Question3.wav'
whisper_result3=whisper_model.transcribe(enquiry3_path, fp16=False)
portfolio_qn3=whisper_result3['text']
print(f" The Extracted text from the Audio: {portfolio_qn3}")

→ The Extracted text from the Audio: The what is the current price of Nvidia stock?

✓ 6s [12] enquiry4_path='/content/Question4.wav'
whisper_result4=whisper_model.transcribe(enquiry4_path, fp16=False)
portfolio_qn4=whisper_result4['text']
print(f" The Extracted text from the Audio: {portfolio_qn4}")

→ The Extracted text from the Audio: What is the best performing stock in my portfolio?

✓ 5s [13] enquiry5_path='/content/Question5.wav'
whisper_result5=whisper_model.transcribe(enquiry5_path, fp16=False)
portfolio_qn5=whisper_result5['text']
print(f" The Extracted text from the Audio: {portfolio_qn5}")

→ The Extracted text from the Audio: Do I have sufficient balance in my brokerage account to purchase five shares of Apple's stock?
```

Figure 14: Screenshot of Text extracted from the audio using Whisper Model

6.6. Streamlit Demo- Module 6

To demonstrate the application, a sample Stream lit app was built to ask the app a sample question about the Portfolio, and the responses were recorded, as shown in the figure.

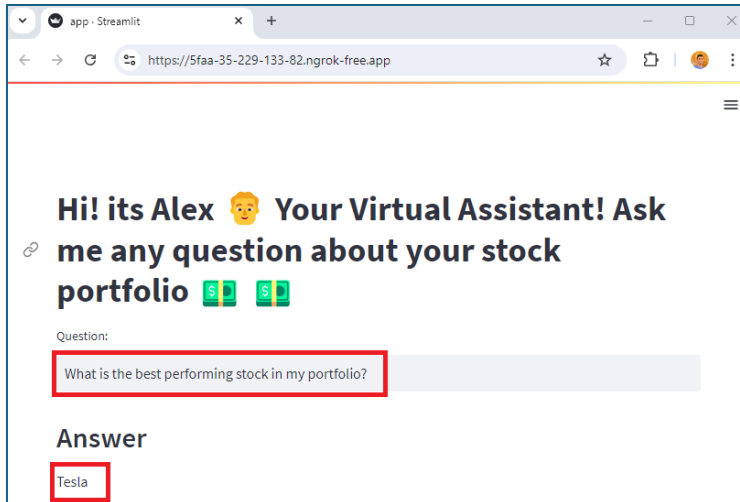


Figure 15: Screenshot of Streamlit App Demo

6.6.1. Validations of the Model Outcome:

1. The outcome of the Voice recognition model was validated by verifying the audio label or listening to the audio.
2. The results returned by the Virtual assistant were validated by executing the queries in the database, as shown in the figure below, which validates the results returned by the Streamlit app.

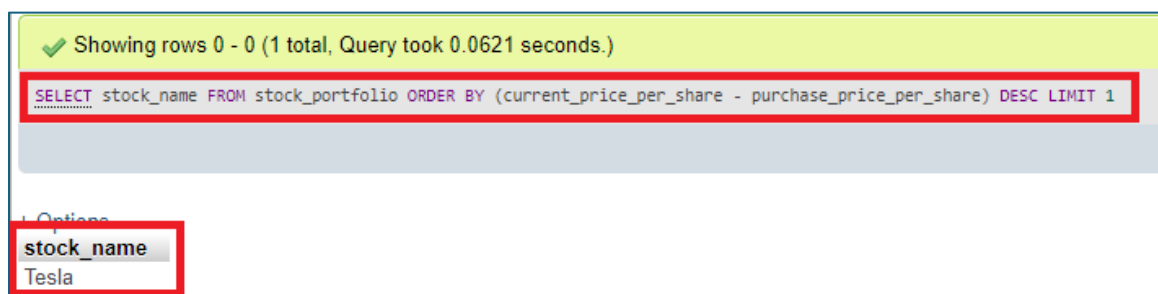


Figure 16: Screenshot of Validation of results by executing the SQL Query

6.7. Research Questions:

Ten Research Questions that Audiences may ask about this Project are:

1. What benefits can the model bring to the table that the existing products on the market lack? How can this model benefit older adults or people with disabilities?

Answer: Two major problems that older adults or People with disabilities face while using Trading or banking apps are difficulty with technology while navigating through the options on the App and the other being difficulty remembering or typing the user ID and password. Both these problems are addressed by voice recognition, which authenticates the user by analyzing the voice, and the virtual assistant built in this project can answer all the questions relating to the account with simple English words.

2. What sectors or services can benefit from the model?

Answer: Though the model was built for Trading platforms, it can be extended to various sectors such as Banking, Utilities, healthcare, etc. Voice recognition, such as biometric authentication, can completely replace the traditional user ID and password authentication methods, thus not requiring the remembering of credentials.

3. What languages can the model understand?

Answer: Currently, the model can only understand English, though expanding to other languages is possible. However, lip reading can take a lot of work to train with other languages as it requires a vast corpus of datasets from various languages to train the model.

4. Is this model a replacement for the authentication methods that are already available?

Answer: Face ID authentication is already available on many smartphones, has wholly redefined biometric authentication, and has replaced the traditional methods of entering passwords each time. However, older adults who are not very familiar with the technology or people with disabilities, such as the Visually impaired, may not find the Face ID authentication very favorable. Though this model does not intend to replace the Face ID authentication, it supplements the existing methods. Also, the Face ID authentication may not work correctly when the customer's face is covered or while wearing a mask, etc. Voice Recognition does not have any of these limitations.

5. How is Lip Reading technology different from Voice Recognition?

Answer: Unlike Voice Recognition, where the customer must speak loudly and clearly, lip reading understands the movements of the mouth and lips and interprets spoken words. There may be situations where the customer cannot speak loudly such as reading out the secret Passphrase in public or in a loud background. Though they have an option to enter the passphrase, the project intends to make the customer's experience easier without having to type or enter anything manually.

6. How can Lip reading help that Face ID authentication cannot?

Answer: Face ID authentication is a more secure mode of authentication than lip reading; however, it comes with a few limitations, such as the visually impaired people not being able to use the face ID, difficulty setting it up, etc.

7. How is the Virtual Assistant built in this project different from a traditional Chatbot?

Answer: The traditional chatbots are usually trained to answer questions from an already pre-trained set of questions, such as FAQs, and they struggle or may not provide satisfactory answers if asked anything outside of this context. The Virtual assistant built in this project with the combination of LLM and Langchain, can answer questions specific to the customer's account, such as the account balance, finding the best-performing stock, etc, which the traditional chatbots cannot.

8. How often will the model retry to authenticate the customer before exhausting the possibilities?

Answer: Though the retry functionality is not included in the model, it is an excellent addition before production implementation.

9. What happens if the Model cannot authenticate the customer? Will there be an option for the customer to contact support?

Answer: If the model has exhausted the allowed number of retries or cannot authenticate the customer, it must redirect to talk to customer support or provide other options, such as entering credentials manually.

10. What are the financial benefits that the Model can provide?

Answer: Many Older adults invest their money in traditional options such as CDs, Bonds, High-interest Savings accounts, etc. They may also seek the assistance of a financial advisor, who usually invests in the stock market for them and charges a fee. As many older adults in their retirement age have more money saved up than an average young adult,

there is a lot of potential if they feel confident using the Trading platform to trade on their own, with the added benefit of skipping the brokerage or commission fees.

11. What are the future enhancements to the Model?

Answer: The model can be enhanced further by reading out the responses returned by the virtual assistant and making it an interactive conversation. Currently, the model can only support English; expanding it to other languages opens many opportunities.

12. What are the drawbacks of the model?

Answer: As Voice Recognition and Passphrase authentication are the main authentication modes, they can be easily exploited by bad artists who can mimic the voices. Hence, additional modes of Protection must be enabled to alert the customer if any unusual activity is observed in the account to protect the customer.

7. Conclusion:

The accuracy of the Voice Recognition model was about 86% on the test datasets, which indicates that there is still some scope for further improvement, as not recognizing the customer's voice or incorrectly providing access to unauthorized users may lead to unforeseen and ethical challenges. The lip reading was challenging to implement, and the model's accuracy was also much lower, indicating that the model is not ready to be deployed in Production. However, the open-source models, such as Whisper and Google Palm, which were already pre-trained, performed very well and predicted correct responses most of the time.

8. Implementation plan:

Multiple models were created in this project, so they all need to work seamlessly before deploying in Production. For instance, if the lip reading cannot correctly recognize the customer's response, it may lock the customer's account, leading to more frustration. Hence, the models must be trained on a large corpus to attain higher accuracy before production deployment. The Virtual Assistant model must be trained in a wide range of questions the customers can ask before deploying in Production.

Once the results are satisfactory, the models can be integrated into the Production phase. The product can be released for a small group of customers or a region, and based on the responses, it can be released further, or actions can be taken to improve the product further.

Also, the model does not include any retry capabilities, which are essential enhancements before deploying them in production.

9. Assumptions:

- The classification models were built assuming that the features portrayed belong to one of the Target Groups and are independent.
- Also, the data was assumed to have no Multicollinearity between the independent features.
- The model was built assuming that most customers use smartphones and Trading apps on their cell phones or the web.
- The model was built assuming that most customers are familiar with the English language and can speak and say the words clearly, as the model performance will be significantly impacted if the words are not spoken clearly.

10. Limitations:

The Models built in this project have various limitations:

- The project does not include a module to record the customer's voice to train the model.
The models were trained using the data available on the internet. In a real-world scenario, the customer must be prompted to repeat the numbers from zero through nine a few times and augment the data to train the model.
- The voice recognition model was trained with the voices of a few young male actors. This could be a significant limitation as the data set needs to include the wide range of customers that the real world represents.
- The SQL Database chain was developed with two tables with just a few sample rows in the table. However, in the real world, there may be many tables with thousands or millions of records in each. Hence, the database performance can become a bottleneck.
- The voice recognition model may not work as expected if the customer is sick and sounds different or in a noisy background.

11. Challenges:

- Older adults may find it challenging to learn and adapt to Technology. Hence, if the Virtual Assistant cannot help the customer, they should be prompted to talk to a Human if they need additional assistance. Many older adults grew up walking to the physical banks and interacting with people, so expecting them to change their way of banking or trading can be challenging for some to adapt.
- Lip reading can be challenging in many situations, as some people may need to speak more clearly or make enough lip movements for the Model to pick the right words. If the Lip

reading or Speaker recognition system cannot assist the customer, they must be provided with an option to call customer support to complete the transaction.

- The models must be trained with many hours of videos to understand a wide range of word corpus effectively.
- The training of the models is language-specific, so extensive training must be done for each language. Different accents of the same language can lead to additional challenges.

12. Future Uses:

- Though the model was built for Trading platforms, it can be extended to various sectors such as Banking, Utilities, HealthCare etc. Voice recognition, such as biometric authentication, can completely replace the traditional user ID and password authentication methods, thus not requiring the remembering of credentials.
- The virtual assistant integrated with the customer's account and being able to answer specific questions relating to individual accounts can unleash a wide range of opportunities in many sectors in the future.
- The virtual assistant can be expanded to perform various tasks, such as Buying and selling shares, apart from answering the basic questions demonstrated in this project.

13. Recommendations:

- The voice recognition model was trained with the voices of young male speakers. As the model is mainly intended for older adults and people with disabilities, it is recommended that the model be trained to use voices representing the general population.

- It is recommended that the Lip-reading model be trained on a vast corpus of datasets to attain higher accuracy before it can be deployed in Production.
- The model was built for English-speaking customers. It is recommended that the model be expanded to other languages.
- It is also recommended that retry capabilities be included so that the model does not quit or deny access to the customer just after one iteration.

14. Ethical Implications:

1. Though the Project aims to provide an enhanced mobile trading and banking experience, it does not intend to entirely replace bankers or financial advisors in physical locations.

Though the corporations may intend to replace human Financial Advisors with Artificial Intelligence solutions, thus reducing costs, it will have many ethical implications, such as Loss of jobs and financial instabilities.

2. Older adults and people with disabilities are the most vulnerable parts of society in terms of financial exploitation. With the constant rise of scammers popping like weeds from everywhere that mainly target seniors and people with disabilities, banks and brokerages should implement more secure measures to protect their customers and alert and try to stop the scams before it is too late. Gravitating more toward automated and AI solutions makes this an even bigger priority for the Brokerages and banks to take more seriously.
3. Customers should always be provided with an option to contact customer support and talk to a person without going through a long wait time or a series of checks to see if they have already authenticated through the mobile app.

15. References:

Faverio, M. (2022, January 13). *The share of those 65 and older tech users has grown in the past decade*. Pew Research Center; Pew Research Center. <https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/>

How—and why—banks can better serve people with disabilities. (n.d.). Deloitte Insights. <https://www2.deloitte.com/us/en/insights/industry/financial-services/accessible-banking-for-disabled.html>

Bennett, R. (2024, February 29). *The Average Savings Account Balance In The U.S.* | Bankrate.com. Bankrate. <https://www.bankrate.com/banking/savings/savings-account-average-balance/#average-savings-by-age>

McEachran, R. (2021, March 22). *How to design banking apps that work for all ages*. Raconteur. <https://www.raconteur.net/finance/banking-app-ages>

The GRID audiovisual sentence corpus. (n.d.). Spandh.dcs.shef.ac.uk. Retrieved July 27, 2024, from <https://spandh.dcs.shef.ac.uk/gridcorpus/>

free-spoken-digit-dataset/recordings at master · Jakobovski/free-spoken-digit-dataset. (n.d.). GitHub. Retrieved July 27, 2024, from <https://github.com/Jakobovski/free-spoken-digit-dataset/tree/master/recordings>

Sugandhi, A. (2023, March 7). *A Guide to Long Short Term Memory (LSTM) Networks*. Www.knowledgehut.com. <https://www.knowledgehut.com/blog/web-development/long-short-term-memory>

Wikipedia Contributors. (2024, June 30). PaLM. Wikipedia; Wikimedia Foundation.

[https://en.wikipedia.org/wiki/PaLM#:~:text=PaLM%20\(Pathways%20Language%20Model\)%20is](https://en.wikipedia.org/wiki/PaLM#:~:text=PaLM%20(Pathways%20Language%20Model)%20is)

Whisper (speech recognition system). (2024, February 24). Wikipedia.

[https://en.wikipedia.org/wiki/Whisper_\(speech_recognition_system\)](https://en.wikipedia.org/wiki/Whisper_(speech_recognition_system))

Appendix

Langchain

Langchain is an open-source framework that lets developers combine the power of artificial intelligence and large language models with external data sources such as their datasets. This can be leveraged in Python, JavaScript, or TypeScript. Langchain overcomes some of ChatGPT's limitations. When ChatGPT came out, it could only answer questions as of 2021. In contrast, Langchain was able to overcome this by accessing the recent dataset and working without any data limitations.