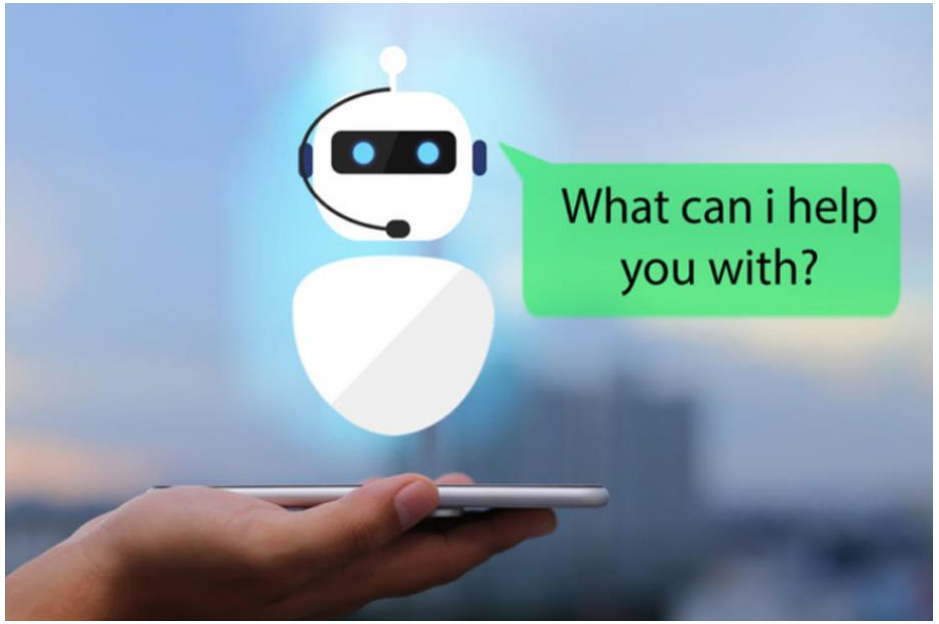
Building a Multimedia Chatbot for AI Testing with Built-In Measurable Intelligence Testability

DATA 298B – Final Demo Project Presentation

Professor: Dr. Zeyu Jerry Gao
San Jose State University

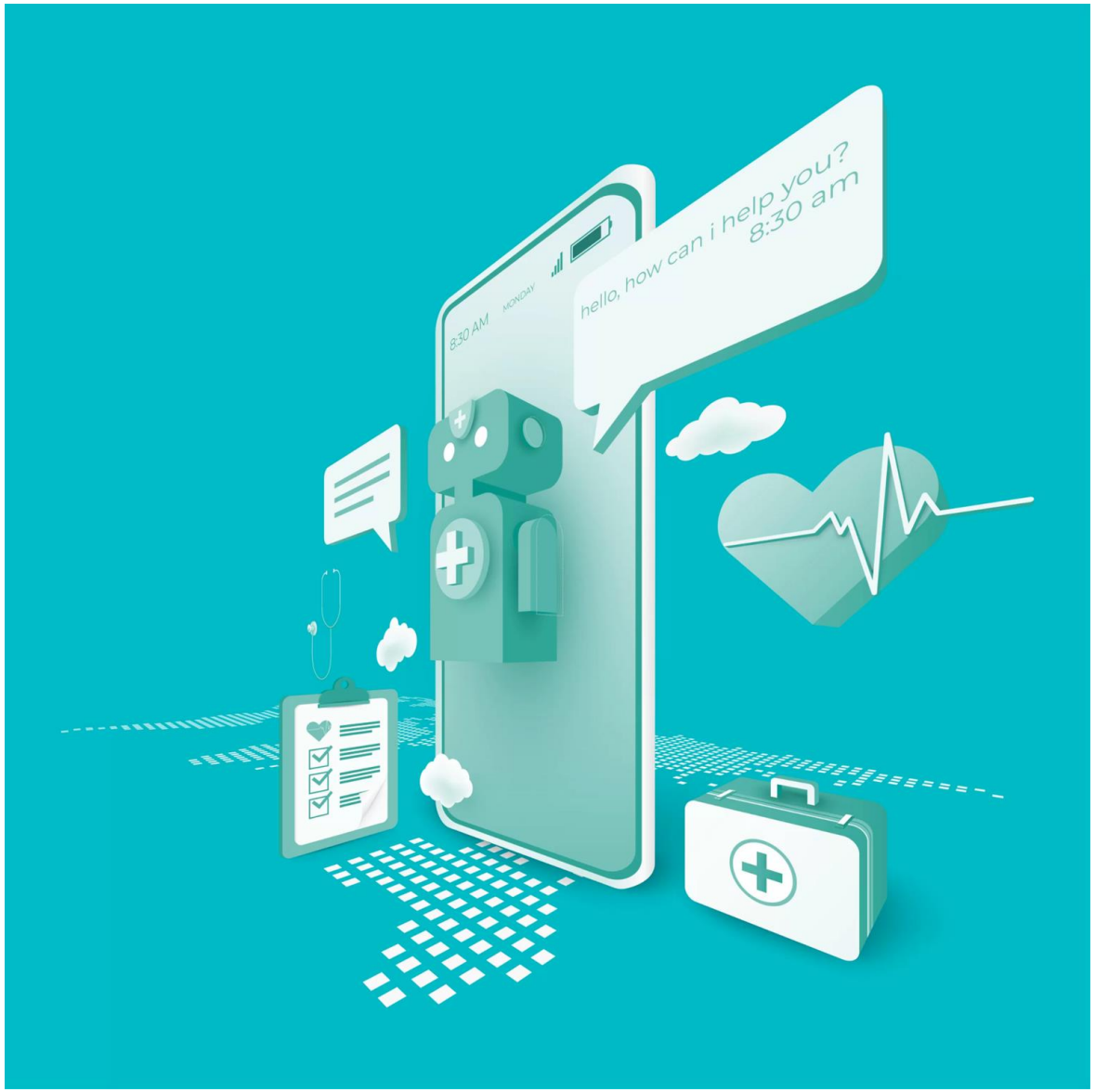
Agenda

- Project Background
- Market Analysis
- Goal & Motivation
- Technology Survey
- Literature Review
- Data Preparation
- Proposed ML Models
- Machine Learning Results
- Web System Design and Development
- System Implementation
- UI Demo



Project Background

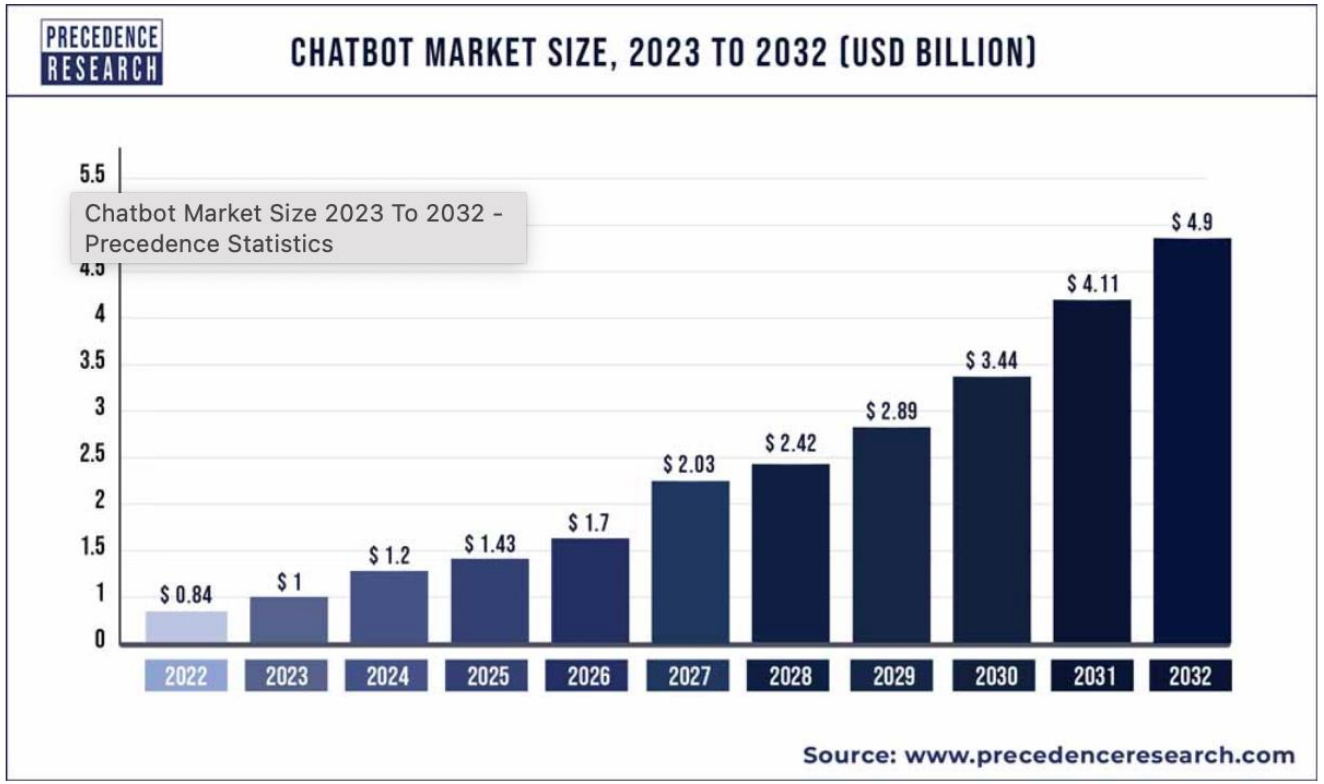
- Introduce a chatbot framework that can offer precise and dependable responses tailored to specific needs.
- Identify relevant answers for customers who may not possess specific and accurate keywords related to the topic.
- Suggest a chatbot that can discern domain-specific information within extended conversations.
- Recommend a chatbot that can assess its own performance following each interaction.
- Outline a standardized assessment approach for measuring the intelligence of chatbots.



Market Analysis

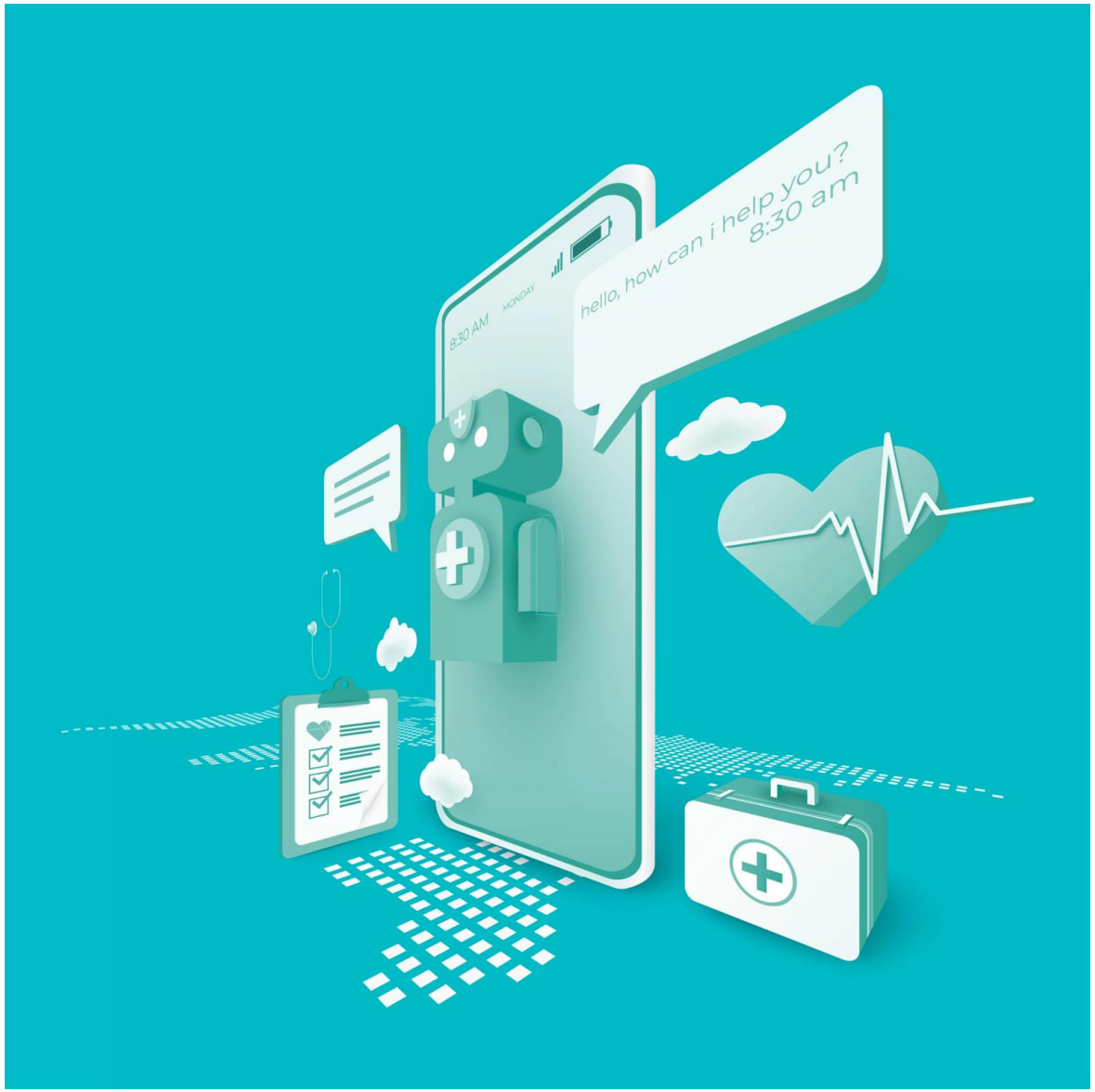
- The AI chatbot market is projected to reach $ 4.9 Billion with an expected growth rate of 28.3% from 2022 to 2032

- AI chatbots are a cost-effective solution for businesses, proven to reduce 30% spending on customer support



Goal & Motivation

- Our goal is to make AI testing tools more user-friendly and accessible to a wider audience
- Build a chatbot that can interact in the form of text, image, emoticons, and audio
- We aim to motivate and empower users to use AI testing tools with better efficiency



Technology Survey

Chatbot name	Platform	Application	Base Model	Type	Performance
Watson Conversation Service by IBM	IBM	Healthcare, Finance, Legal, Retail, Fantasy Football	LSTM, CNN, Transformer	Retrieval Based	Accuracy: 70-80%
DialogFlow	Google	Customer Service, E-commerce, Education	RNN, Transformer, BERT	Rule based + ML based	Accuracy: 85-95%
RASA NLU	Open source	Customer Service, E-commerce	LSTM, Transformer	Machine Learning	Accuracy: 90%
Microsoft Bot Framework	Microsoft (Slack, Skype, Facebook)	Customer Service, E-commerce, Healthcare	NLP, RNN, LSTM	Machine Learning	Accuracy: 80-90%
Speech-to-Text chatbot	Google API	Healthcare	RNN, LSTM, Transformer	Rule-based + ML based + Knowledge based	F1- score: 88%
Chatbot Based image classification	LINE	Retail e-Commerce	CNN, RNN, GAN	Rule based + DL based	Accuracy: 80-90%

Data Preparation Process

Data Collection Process for Text Data

Self Collected Dataset: Text-based question-and-answer (Q&A) pairs for each section and functionality of the AI testing tool

Existing Dataset: Amazon dataset on Electronics along with original data to increase diversity and generalization

JSON-formatted Amazon data contained 38000 records which was converted into structured format

Text Q&A Data Sample

Self Collected Dataset for AI Testing Tool	
Source	Primary Data
Size	25 KB
Format	CSV
Number of instances	117
Number of Features	2
Sample1	Ques: How to signup in the AI testing tool? Ans: Click on the signup button

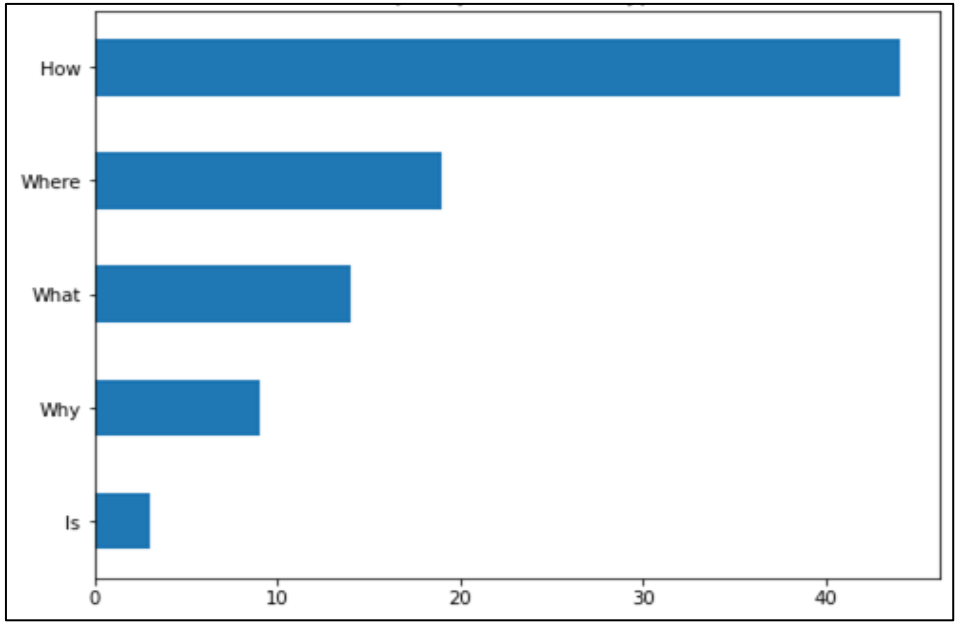
Amazon Dataset on Electronics	
Source	Primary Data
Size	58 MB
Format	CSV
Number of instances	153272
Number of Features	2
Sample1	Ques: is it user friendly? Ans: It is pretty user friendly/ self explanatory.

Text Q&A Data Sample

Self Collected Dataset Sample

Questions	Answers
How to signup in the AI testing tool?	Click on the sign up button
How to sign up for AI testing tool	Click on the sign up button as shown in the image
Why I am getting an error while login?	Either the password or username is incorrect. Enter the correct credentials.
Why I am not able to login to AI testing tool?	Username and Password is incorrect. Please enter the coorect credentials
How to create a new project?	Go on to Dashboard and click on add project. You can start creating new project.
How to look at previous created function?	Click on project management module to view the previously created functions
How to add a project?	Go on to Dashboard and add project
Where can you start creating a new project?	Go on to the Dashboard to create project
How to look at previous project?	Go to the Dashboard and click on the summary section
How to check all uploaded projects?	specific project name in the filter
How to search for a project?	Go to the Project Management tab and then type in the project name in filter

Frequency of Question Types



Augmented Text Q&A Data Sample

Augmented Question(Positive) Using BERT and NLP Aug

Original Question	Augmented Question (Positive)
How to sign up to the AI testing tool?	how to compete in the ai integration process?'
How to sign up to the AI testing tool?	listening to signup in the ai recruitment process?'
How to sign up to the AI testing tool?	how to enroll in the competitive testing program?'
How to sign up to the AI testing tool?	learn to signup in that math testing tool?'
How to sign up to the AI testing tool?	how near signup via another ai testing tool?'
How to sign up to the AI testing tool?	how to use as the project testing tool?'
How to sign up to the AI testing tool?	anybody to trust in any ai testing tool?'
How to sign up to the AI testing tool?	how to survive in the public news tool?'
How to sign up to the AI testing tool?	how long signup before the math testing tool?'
How to sign up to the AI testing tool?	how to work on an ai testing tool?'

Augmented Question(Negative) Using BERT and NLP Aug

Original Question	Augmented Question (Negative)
How to sign up to the AI testing tool?	How to signup in the?'
How to sign up to the AI testing tool?	To signup in the AI testing?'
How to sign up to the AI testing tool?	To in AI testing tool?'
How to sign up to the AI testing tool?	To how signup in the testing tool AI?'
How to sign up to the AI testing tool?	How signup to in AI testing the tool?'
How to sign up to the AI testing tool?	To how in signup AI the testing tool?'
How to sign up to the AI testing tool?	how to provide signup in to the web ai testing tool?'
How to sign up to the AI testing tool?	how to match signup in the the ai testing tools tool?'
How to sign up to the AI testing tool?	How to signup in tool?'
How to sign up to the AI testing tool?	To in the AI testing tool?'

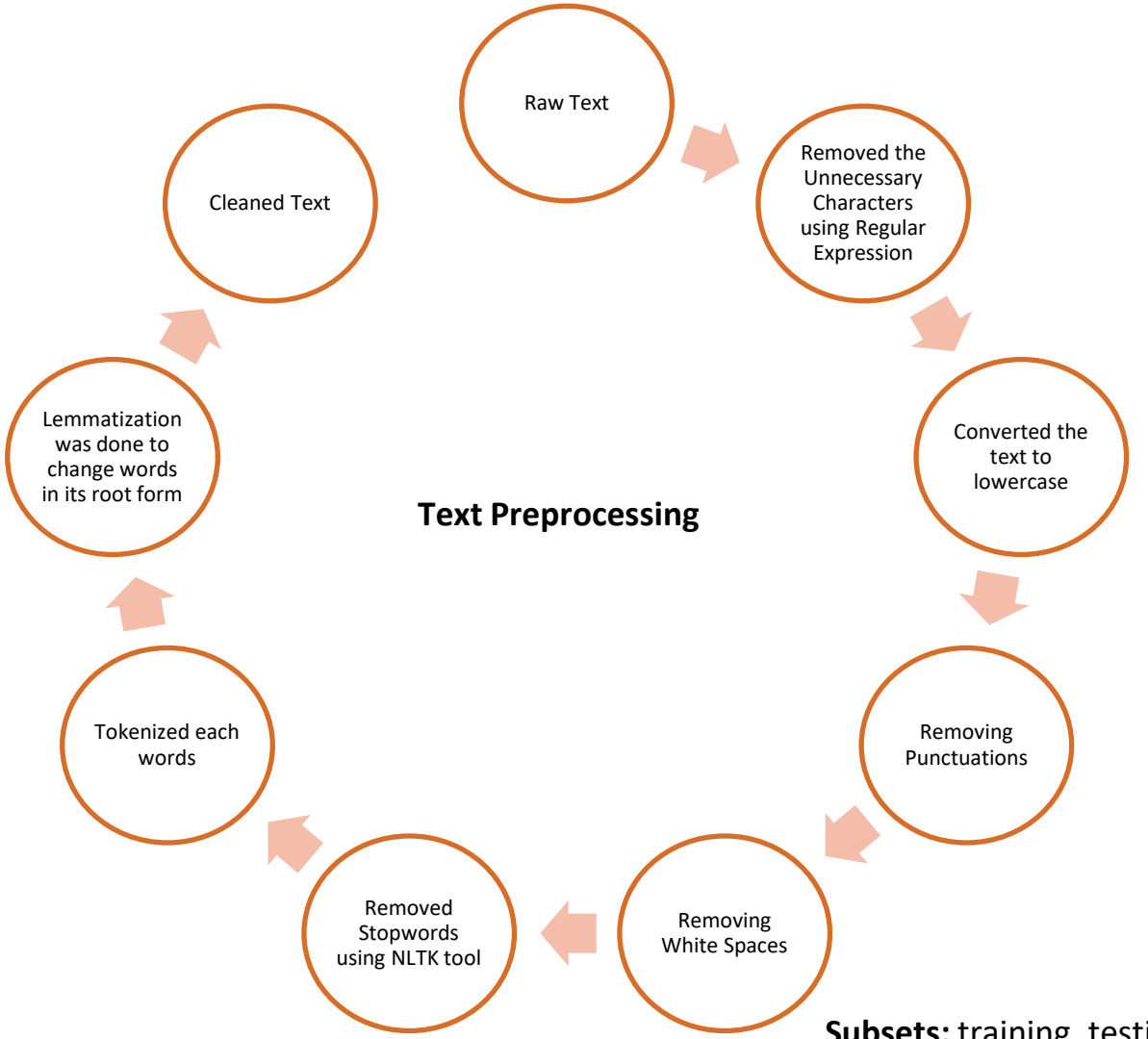
Augmented Question(Positive) Using GPT3 API

Original Questions	Augmented Answer (Positive)
How to signup to the AI testing tool?	How can I sign up for the AI testing tool?
How to signup to the AI testing tool?	What are the steps to sign up for the AI testing tool?
How to signup to the AI testing tool?	To sign up for the AI testing tool, go to the website and click on the "Sign Up" button.
How to signup to the AI testing tool?	How can I sign up for the AI testing tool?
How to signup to the AI testing tool?	How can I sign up for the AI testing tool?
How to signup to the AI testing tool?	What is the process for signing up for the AI testing tool?
How to signup to the AI testing tool?	To sign up for the AI testing tool, go to the website and click on the "Sign Up" button.
How to signup to the AI testing tool?	To sign up for the AI testing tool, go to the website and click on the "Sign Up" button.
How to signup to the AI testing tool?	To signup for the AI testing tool, go to the website and click on the signup button.
How to signup to the AI testing tool?	How can I sign up for the AI testing tool?

Augmented Question(Negative) Using GPT3 API

Original Questions	Augmented Answer (Negative)
How to signup to the AI testing tool?	How to signup in the testing tool?
How to signup to the AI testing tool?	How to signup in AI the testing tool?
How to signup to the AI testing tool?	random_word How to signup in the AI testing tool?
How to signup to the AI testing tool?	How to in the AI testing tool?
How to signup to the AI testing tool?	How to signup tool? the AI testing in
How to signup to the AI testing tool?	How to signup random_word in the AI testing tool?
How to signup to the AI testing tool?	How to signup in the AI tool?
How to signup to the AI testing tool?	How to the in signup AI testing tool?
How to signup to the AI testing tool?	random_word How to signup in the AI testing tool?
How to signup to the AI testing tool?	How to in the AI testing tool?

Data Preprocessing Process for Text Data



Pre-processed Dataset

Raw Text	Cleaned Text
How to signup in the AI testing tool?	signup ai testing tool
How to sign up for AI testing tool	sign ai testing tool
Why I am getting an error while login?	getting error login

Words Embedding

```
Original sentence: How to sign up for AI testing tool
------------
Word index: {'how': 1, 'to': 2, 'sign': 3, 'up': 4, 'for': 5, 'ai': 6, 'testing': 7, 'tool': 8}
------------
Sequences: [[1, 2, 3, 4, 5, 6, 7, 8]]
------------
Padded sequences: [[0 0 1 2 3 4 5 6 7 8]]
------------
Embedded sequence: [[[ 3.2538954e-02  2.3213062e-02  1.7903056e-02 -2.7007019e-02
   -4.0899180e-02]
  [ 3.2538954e-02  2.3213062e-02  1.7903056e-02 -2.7007019e-02
   -4.0899180e-02]
  [-2.8629526e-03  1.8944968e-02  5.2422285e-03 -2.4683034e-02
    2.7922343e-02]
  [-4.2583633e-02  3.3570673e-02 -8.3799362e-03 -2.7657533e-02
   -4.5284998e-02]
  [ 1.3904337e-02  1.0424864e-02  4.9552653e-02  4.1416135e-02
   -3.1513587e-02]
  [-3.7351727e-02 -1.6078569e-02  3.6366593e-02  1.1796571e-02
   -4.0301967e-02]
  [-1.7065503e-02  1.1351358e-02 -2.1128511e-02 -1.5907444e-02
   -7.6293238e-03]
  [ 1.0993253e-02 -7.7252015e-03 -2.8143514e-02 -1.6375054e-02
    2.5767390e-02]
  [-2.1774078e-02  1.1231191e-03 -2.4809336e-02 -3.5895705e-03
   -4.2013370e-02]
  [-9.8299235e-05 -3.8061477e-02  8.2844496e-03 -4.6945501e-02
   -1.5255667e-02]]]
```

Subsets: training, testing, and validation
Splitting the Data: Train : 128K ; Test = 16K, Validation = 16K instances

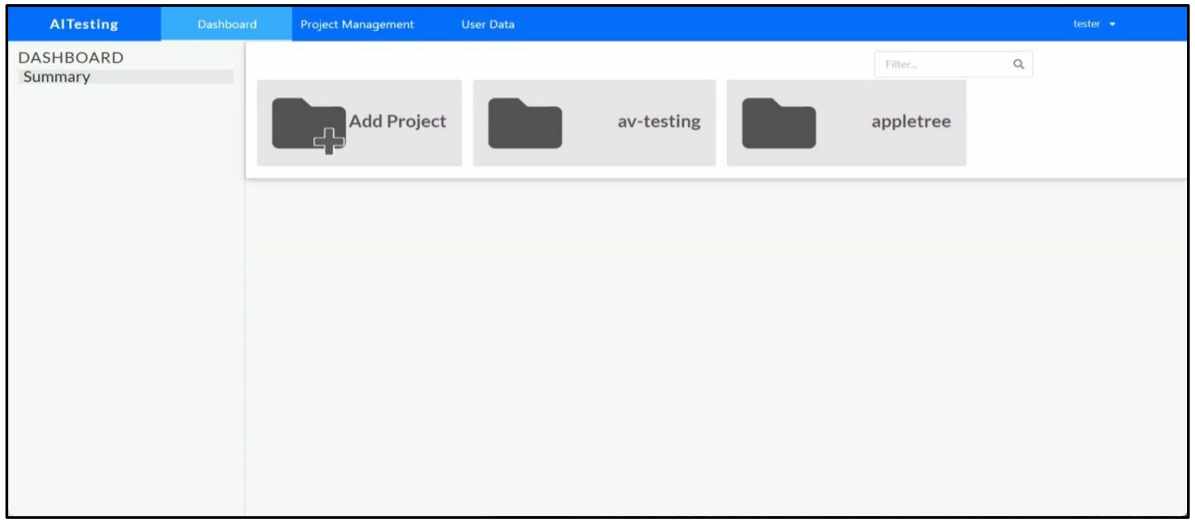
Data Collection – Image Data

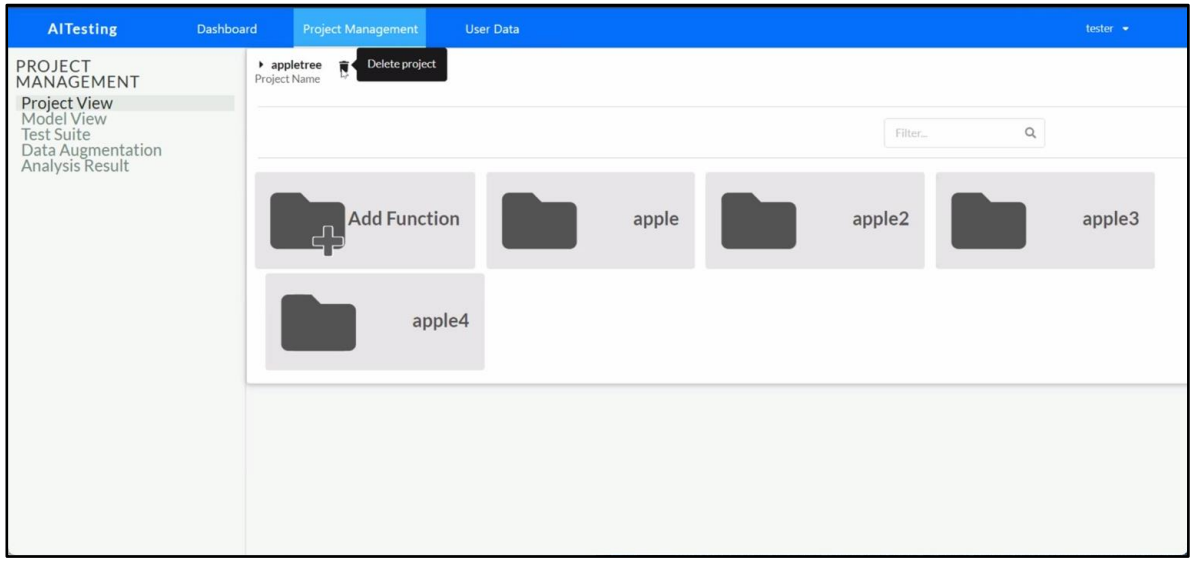
Captured screenshots for various pages of AI testing tool:

- ❖ Project Management Tab
  - Project View
  - Model View
  - Test Suite
  - Data Augmentation
- ❖ Dashboard Tab
  - Summary
- ❖ User Data Tab

Consists of UI components

- Text Buttons
- Icons
- Dropdown boxes
- Checkbox
- Input Fields



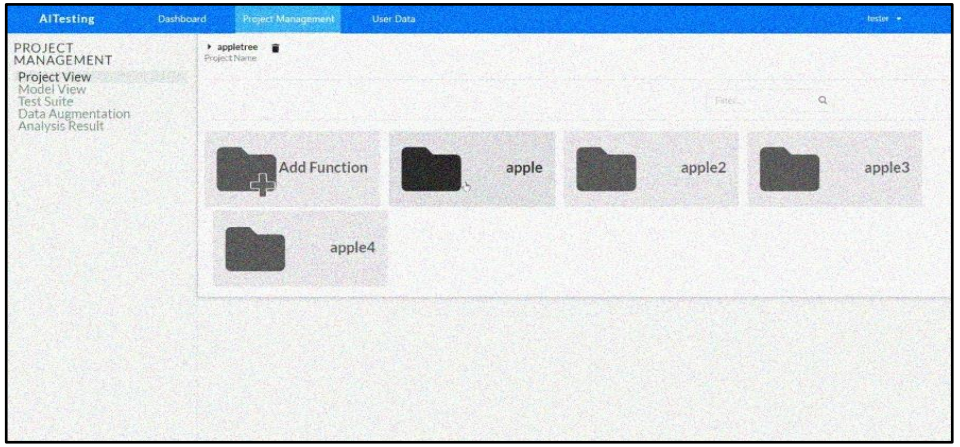


Data Pre-processing - Image Data

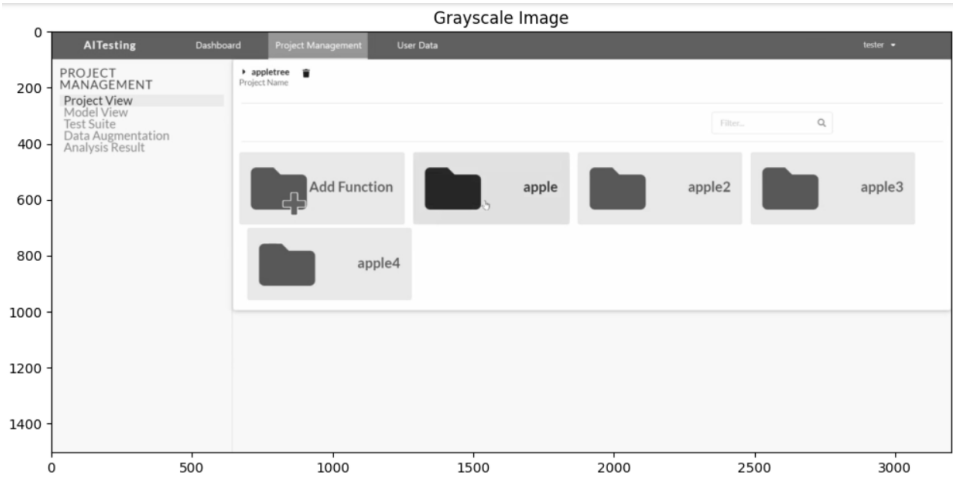
- Data Cleaning: Dropped the Blank images
- Image Resizing: OpenCV
- Image Annotations: LabelImg
  - To segment the images
  - Extract the Region of Interest from the images
  - Identify the UI elements and the relevant text

Data Transformation - Image Data

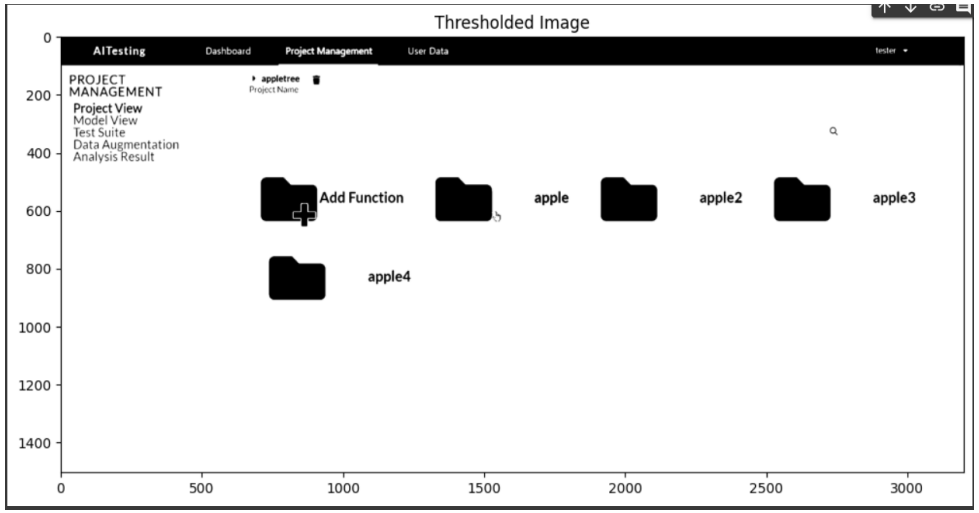
Random Noise Addition



Grey-Scaling



Thresholding - Binarization

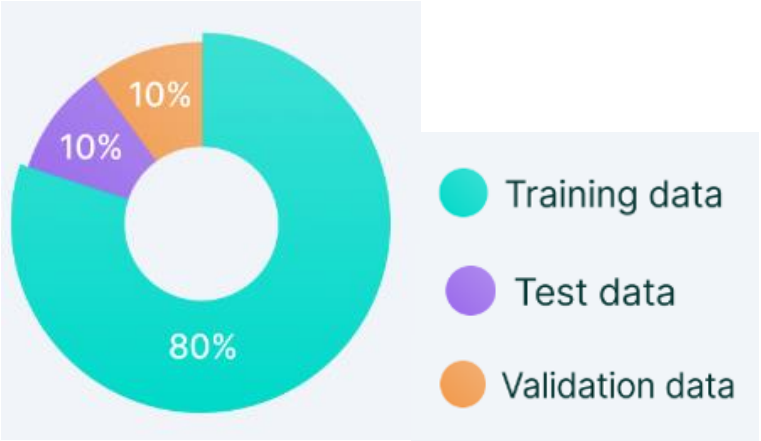


Data Preparation - Image Data

Dataset	Count of Images
AI Testing Tool Images	2400
ICDAR dataset	1500
VINS dataset	4800

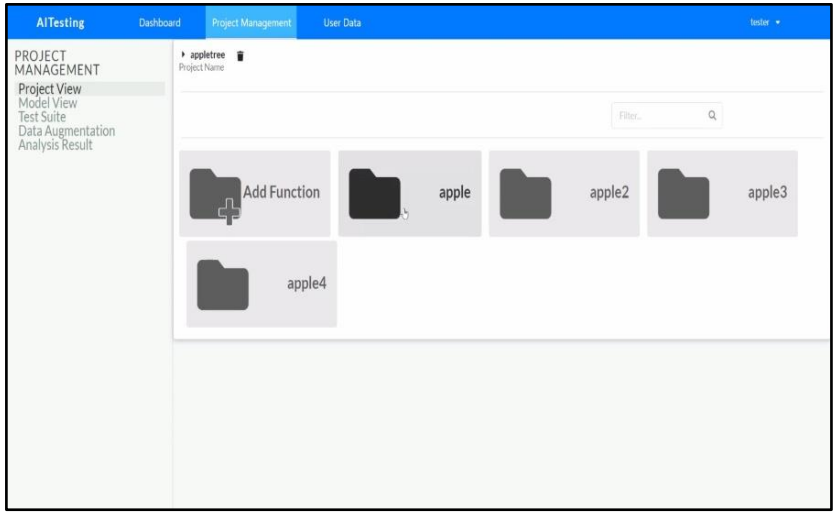
Train

Split Ratio



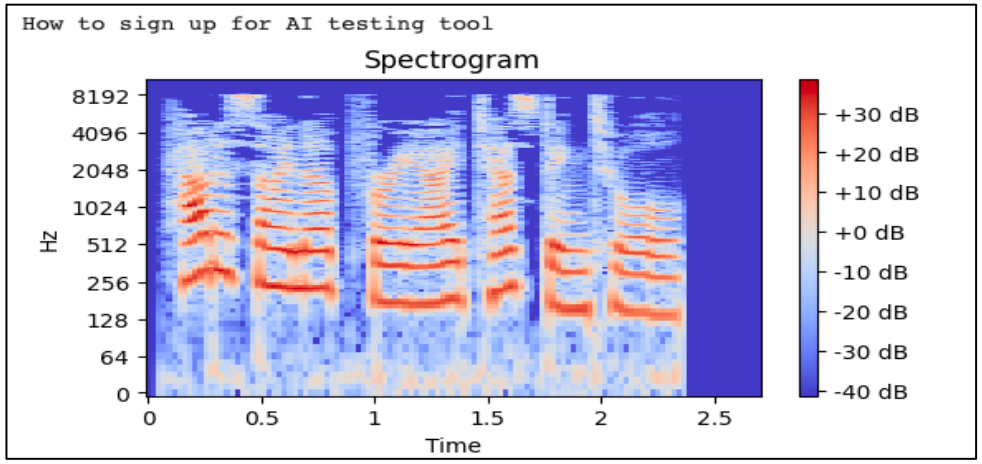
Test

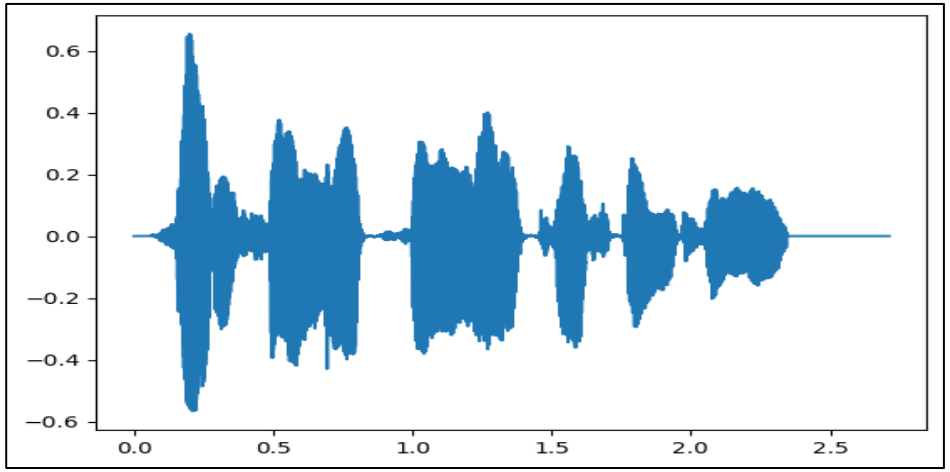
Validate

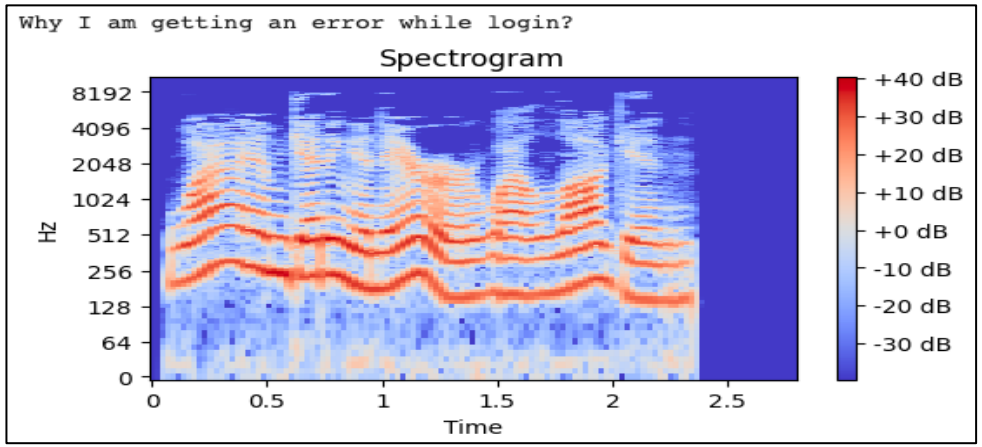


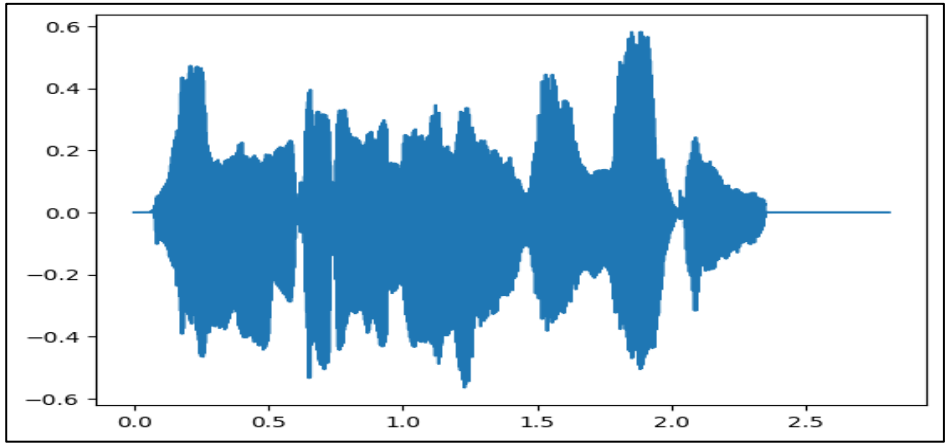
Audio Data Raw Sample

	Questions	Answers	audio
0	How to signup in the AI testing tool?	Click on the sign up button	/content/drive/MyDrive/AI_Testing_Chatbot_DATA...
1	How to sign up for AI testing tool	Click on the sign up button as shown in the image	/content/drive/MyDrive/AI_Testing_Chatbot_DATA...
2	Why I am getting an error while login?	Either the password or username is incorrect. ...	/content/drive/MyDrive/AI_Testing_Chatbot_DATA...
3	Why I am not able to login to AI testing tool?	Username and Password is incorrect. Please ent...	/content/drive/MyDrive/AI_Testing_Chatbot_DATA...



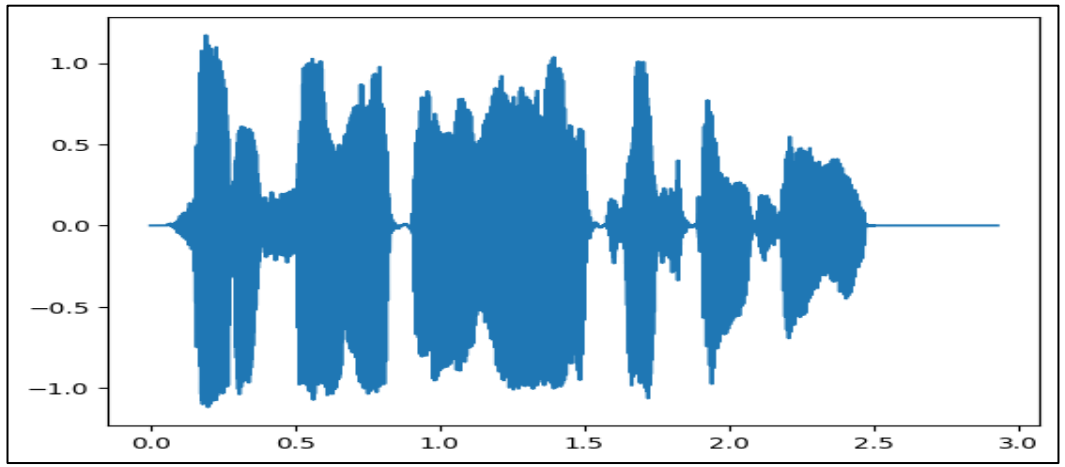






Audio Data Preprocessing

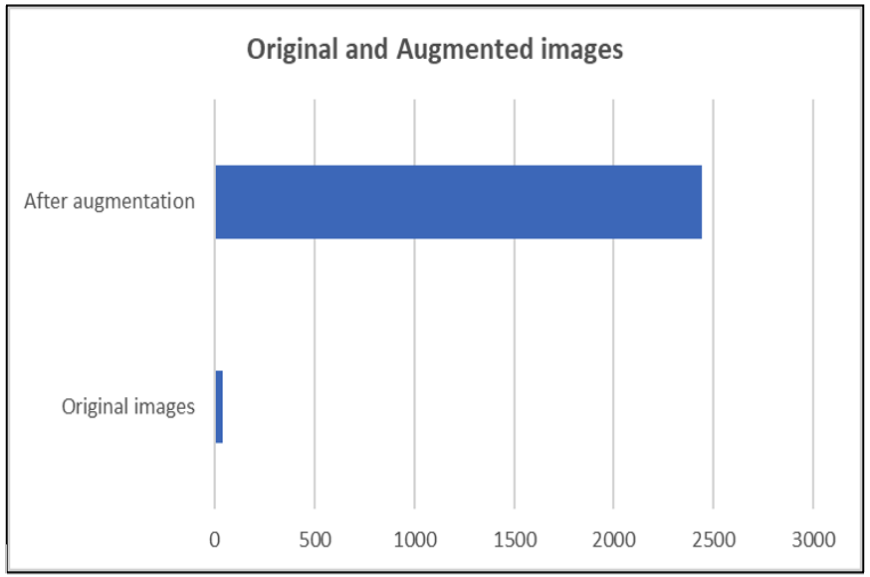
- "transcribe()" processes an audio file using a model to transcribe speech into text, analyzing 30-second windows at a time
- "whisper.detect_language()" detects the language of the audio being transcribed
- "whisper.decode()" provides advanced access to the model's components for customization and improved performance

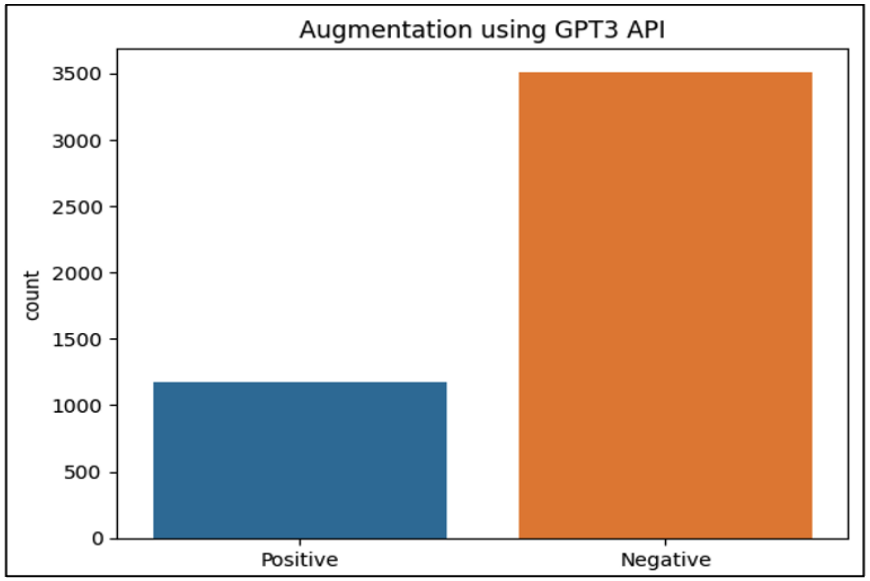


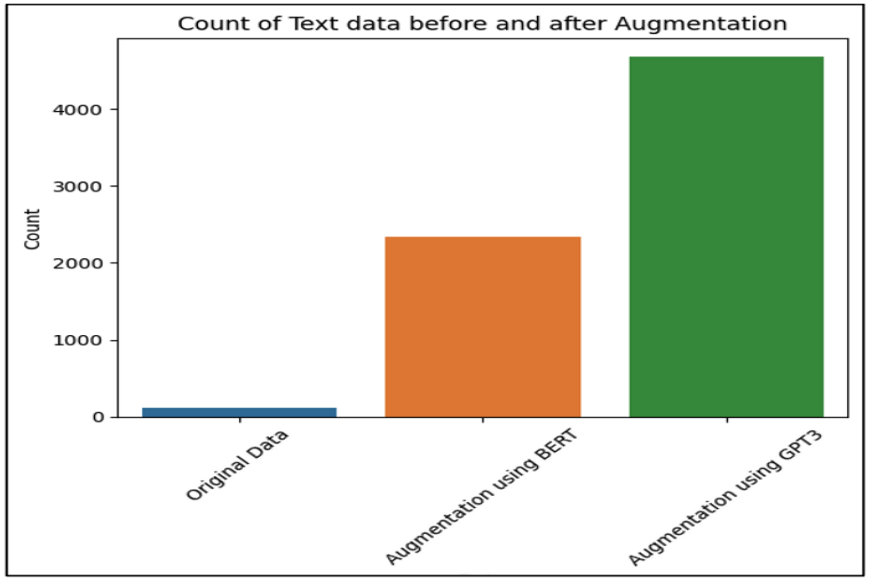
Data Statistics

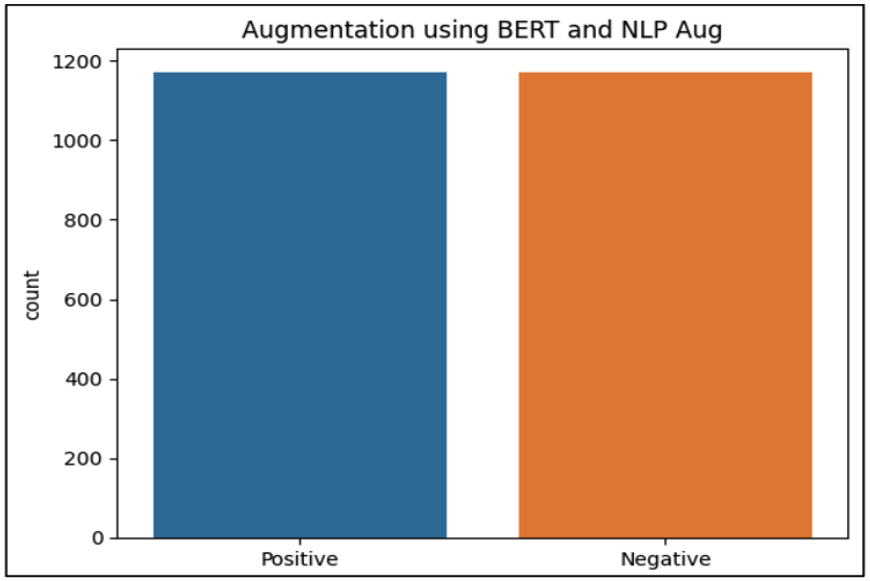
Data Source	Raw Data	After Transformation	Prepared Data		
			Training	Validation	Testing
Text Data	159k	160k	128k	16k	16k
Image data (AI testing screenshots + ICDAR+ VNSI)	6340	8740	6992	874	874
Audio	125	7600	6080	760	760

Contd


Original and Augmented images
After augmentation
Original images
0
500
1000
1500
2000
2500
3000

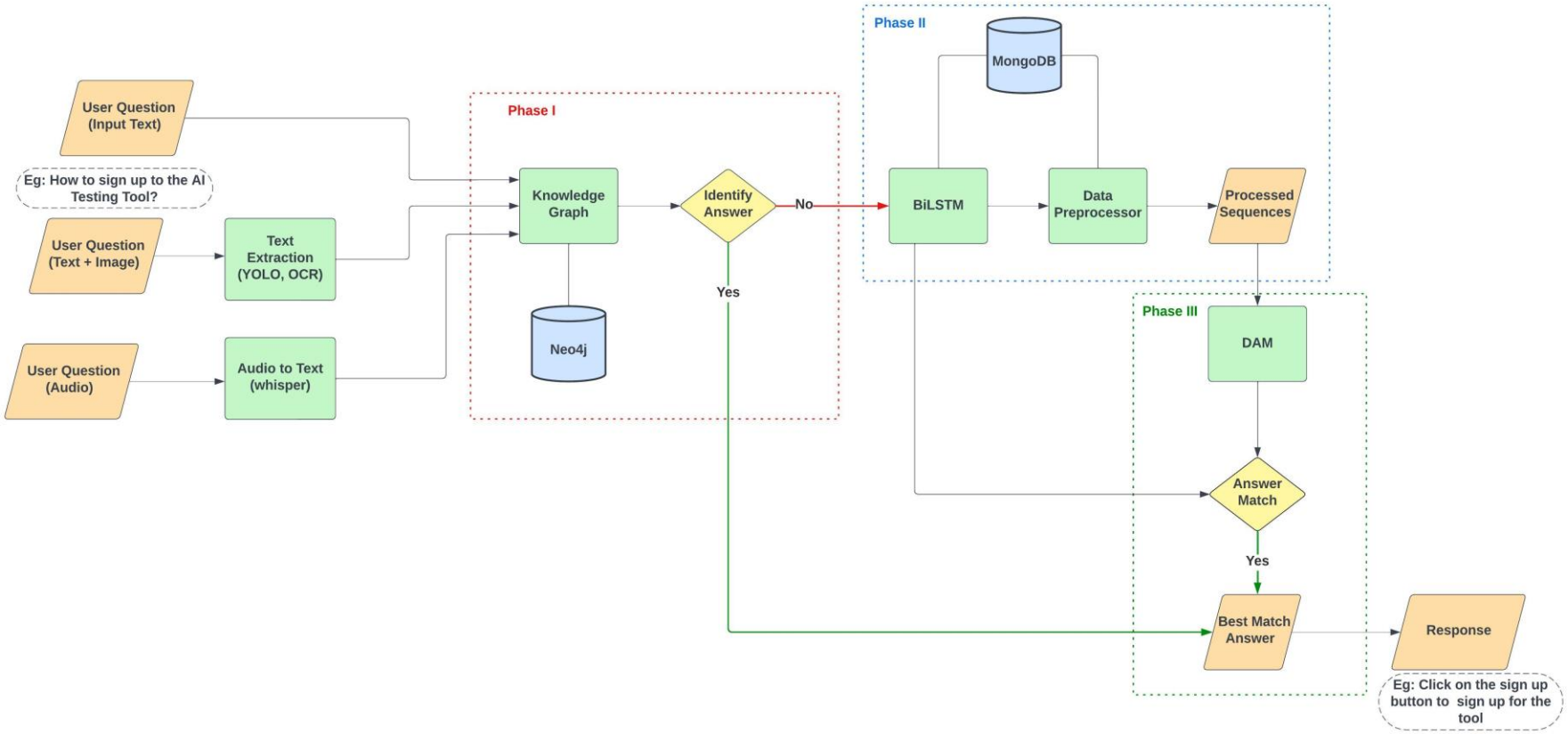

Augmentation using GPT3 API
count
3500
3000
2500
2000
1500
1000
500
0
Positive
Negative


Count of Text data before and after Augmentation
Count
4000
3000
2000
1000
0
Original Data
Augmentation using BERT
Augmentation using GPT3


Augmentation using BERT and NLP Aug
count
1200
1000
800
600
400
200
0
Positive
Negative

Proposed ML Models

Model Overview - Integrated Model



YOLO + OCR
For image input, extract the GUI elements using YOLO v5 model and the relevant text using OCR

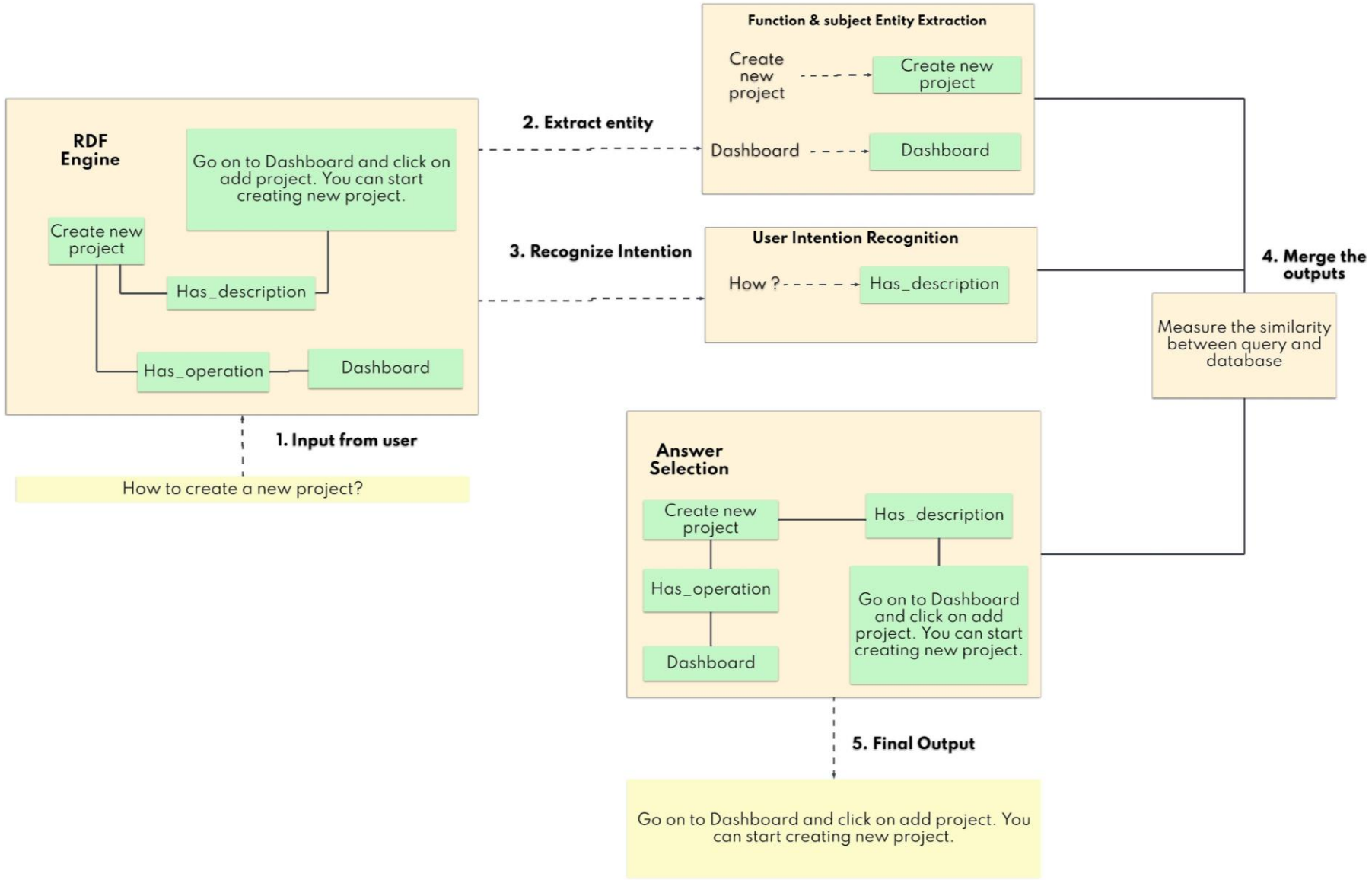
Whisper
For audio input, convert the audio to text data

Knowledge Graph:
Retrieve relevant answer from Neo4j database based on similarity scores between relevant answers

BiLSTM:
Corrects any typos in the original input question and suggest a set of closest matching questions from the database

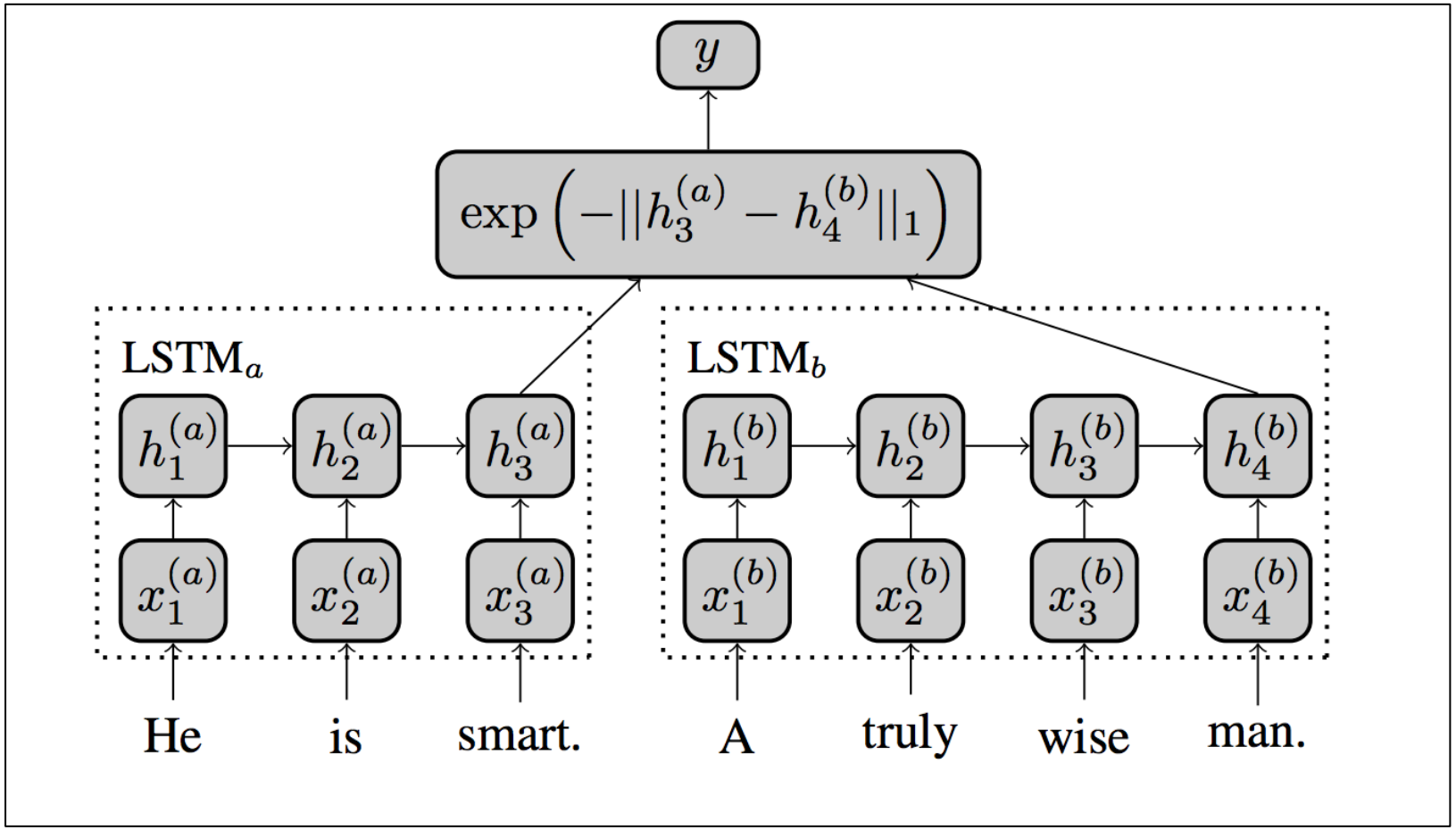
DAM:
Detects the paired answer with the highest match score

Knowledge Graph Model



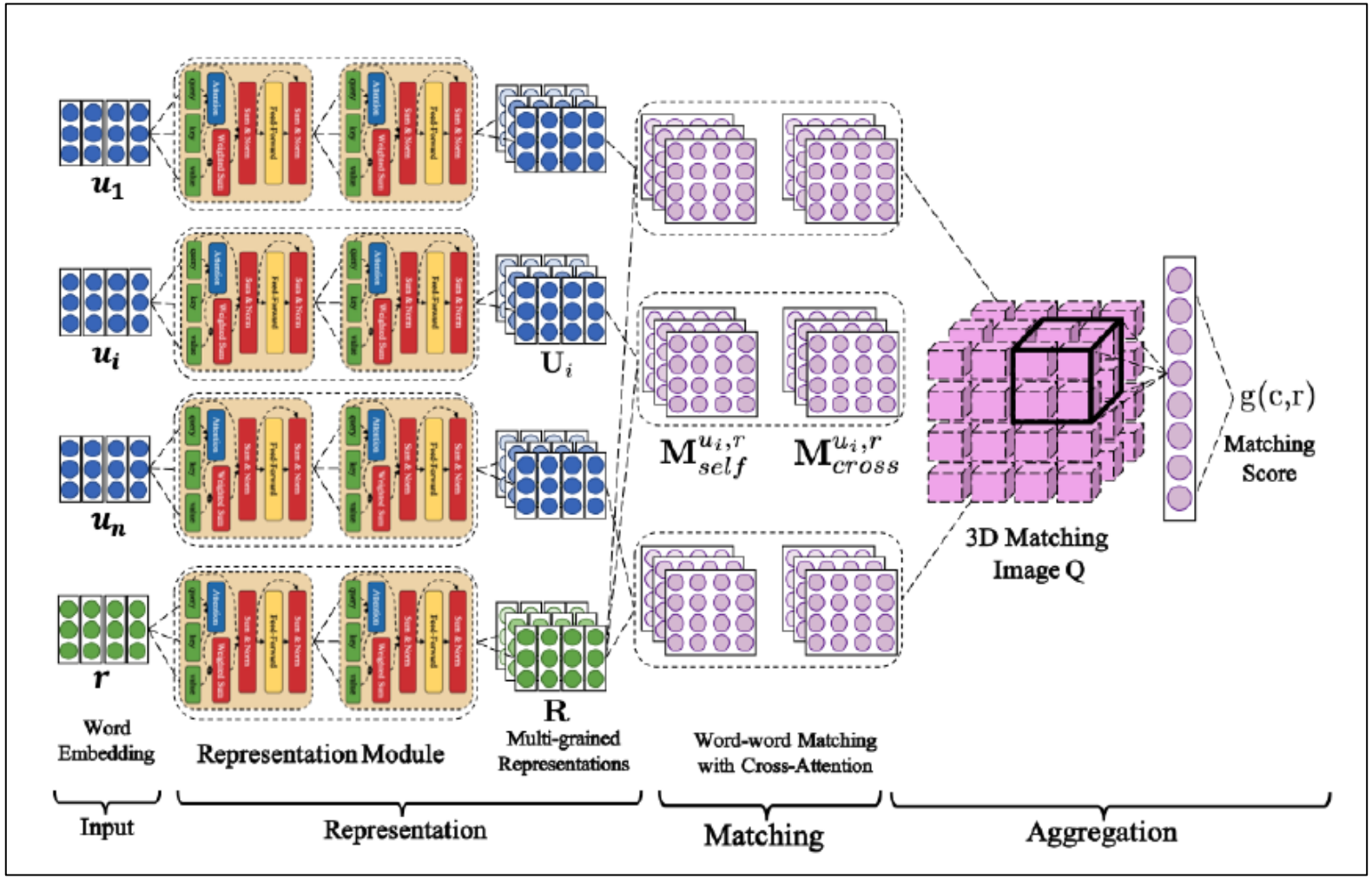
1. The user enters a query.
2. The function and subject entity extractor is used to extract entities.
3. Intention recognition identifies the user's intention.
4. To arrive at an answer, entities and intentions are combined.
5. The user is subsequently given the answer.

Siamese Bi-LSTM model



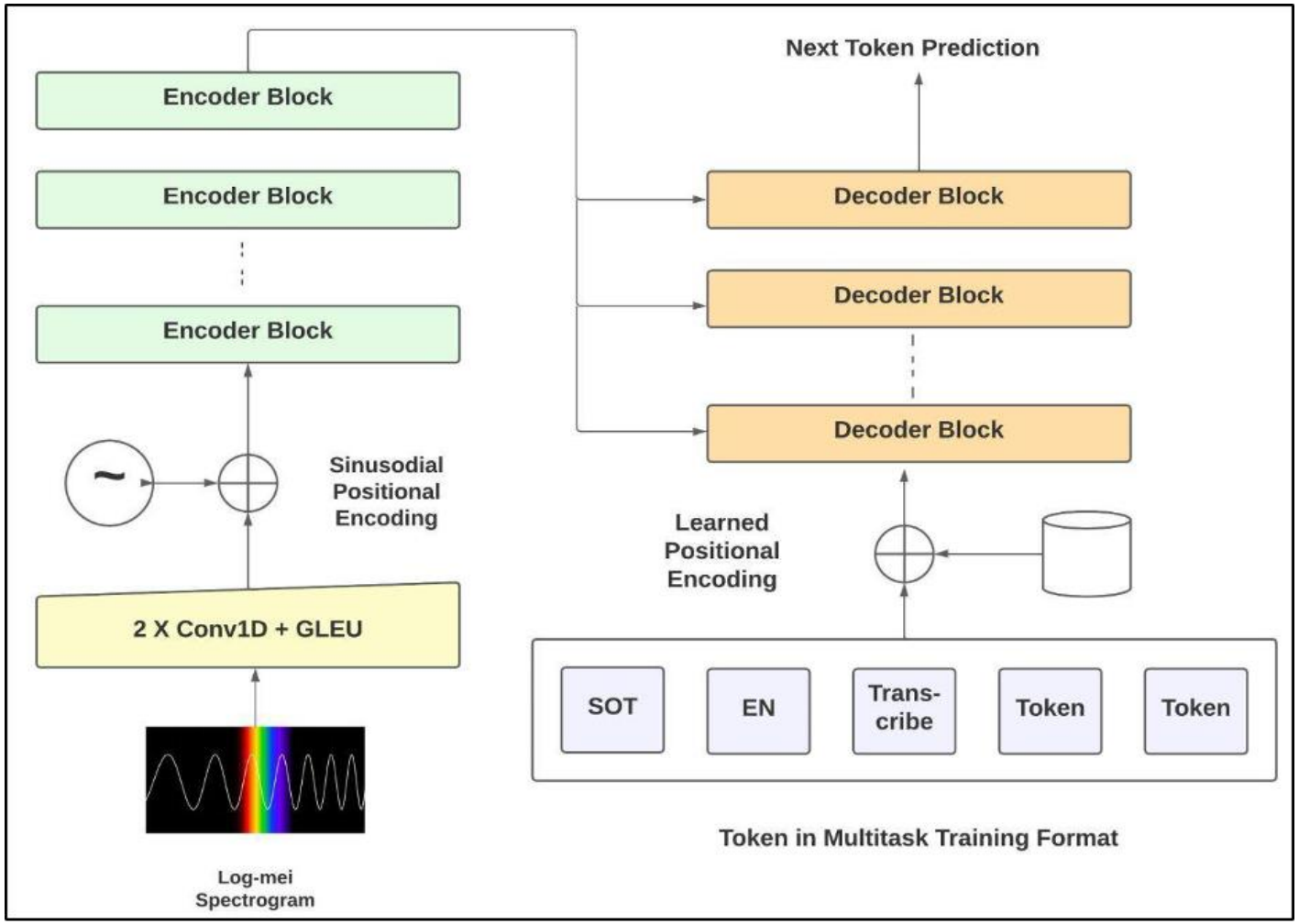
- Designed to determine the similarity between pairs of sentences/questions (Correct question and augmented one's)
- Employs the Manhattan distance formula to compute the connection between the two questions through the implementation of the Siamese architecture

Deep Attention Matching Network



- Captures the semantic links between various conversational elements throughout the process of numerous rounds of conversation
- Creates multiple embedding expressions, enhancing response complexity.
- Mself and Mcross yield six diverse matching matrices for analysis.
- Applied 3D convolution and pooling for final matching score

Audio Input: Whisper Speech-to-Text Model

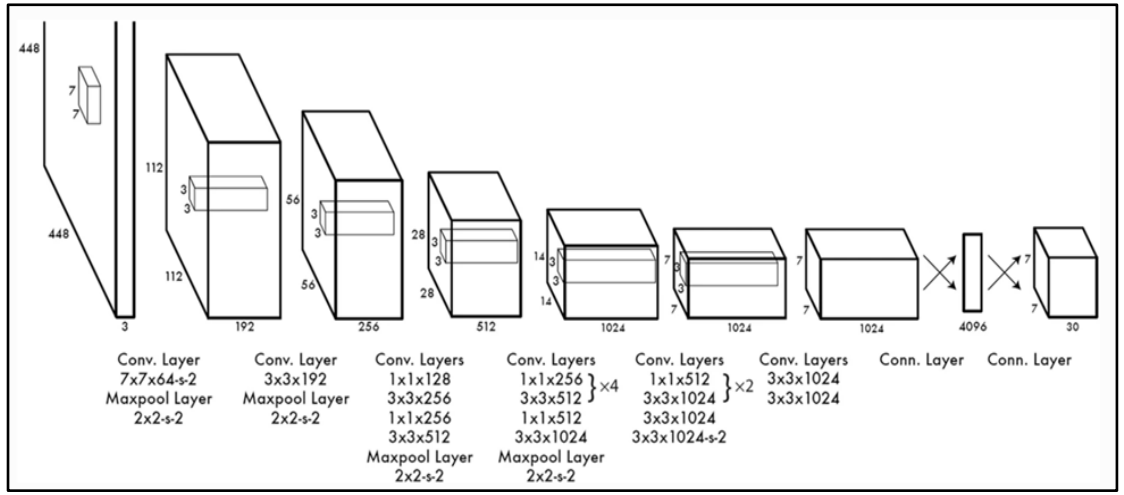


- The Whisper architecture uses an end-to-end technique with a simple encoder-decoder Transformer
- Audio input is divided into 30-second segments and transformed into a log-Mel spectrogram
- The encoder processes the spectrogram, and a decoder is trained to predict labeled text using unique tokens

Image Input: YOLO + OCR

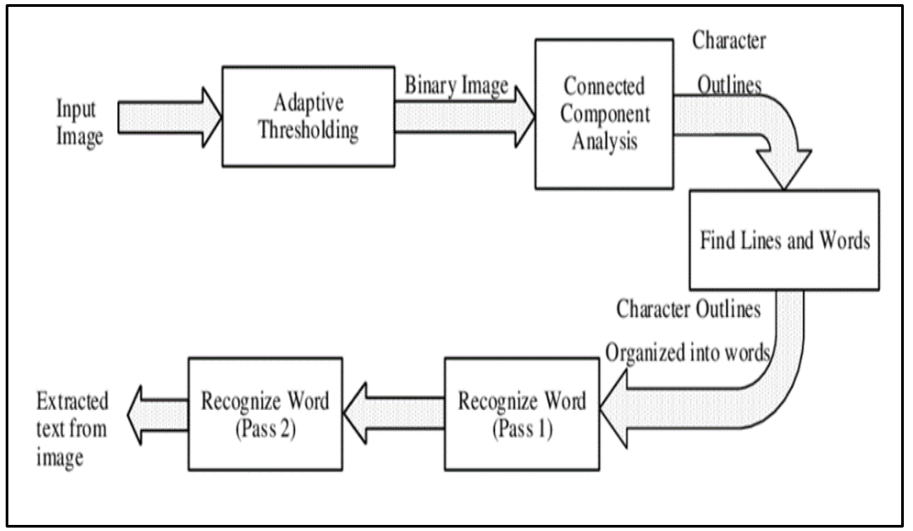
YOLO

- Single-stage object detection algorithm
- Predicts the bounding boxes and class probabilities for multiple objects in an image
- Used for detecting and locating specific objects like
  - Text buttons
  - icons
  - UI elements from AI testing tool screenshots



Tesseract OCR

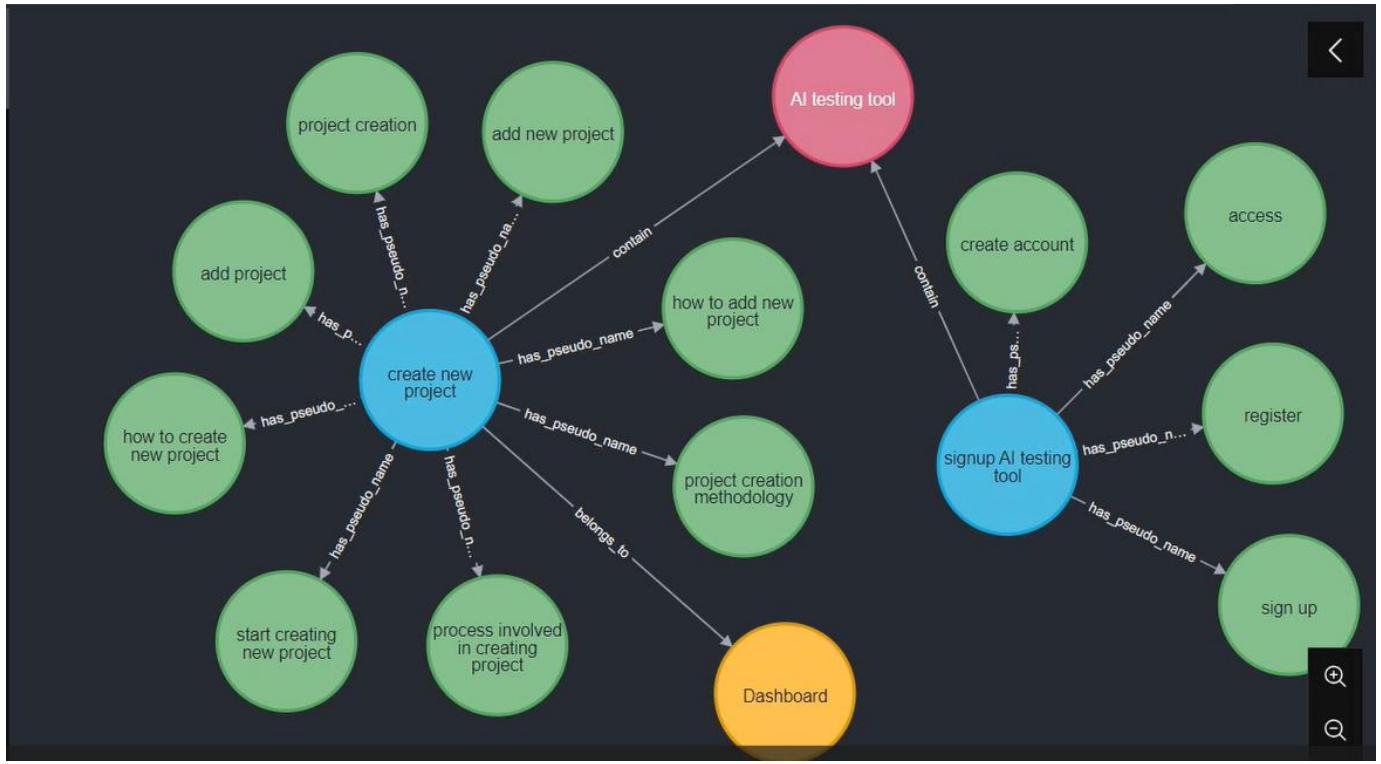
- Text Extraction Technique from images
- Widely used open-source OCR engine
- Uses image processing techniques and deep learning approach (CNN + RNN) for text recognition
- Used for extracting the text from the GUI elements detected from YOLO/Faster-RCNN

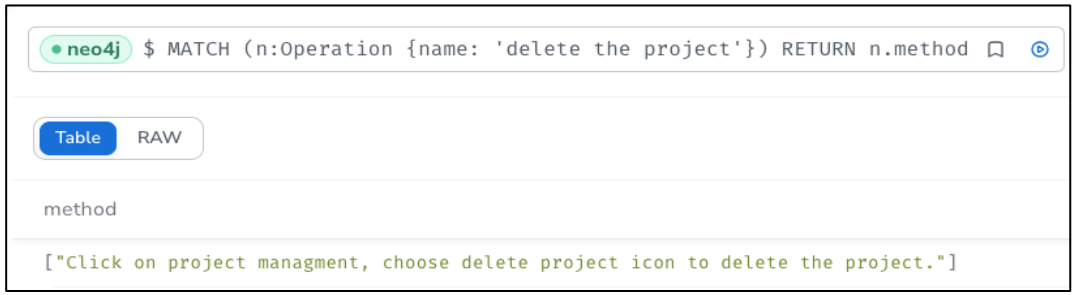


Machine Learning Results

Knowledge Graph

Demo link: https://drive.google.com/file/d/1fphMhW19-bZ2EExd3MKZT3Z0sKtJ0G_s/view?usp=drive_link





- Knowledge graph is designed using Neo4j by defining context and relationships between nodes, retrieving efficient response using Cypher query
- Model returns the most relevant answer by calculating cosine similarity score between user questions and stored responses
- Schema can be accessed by anyone with Neo4j login credentials

Ongoing Machine Learning Results

When knowledge graph doesn't provide relevant answers, Siamese BiLSTM model predicts the best-matched answers from database

Data pre-processor input text into embedded sequences. Paired sequences go through DAM

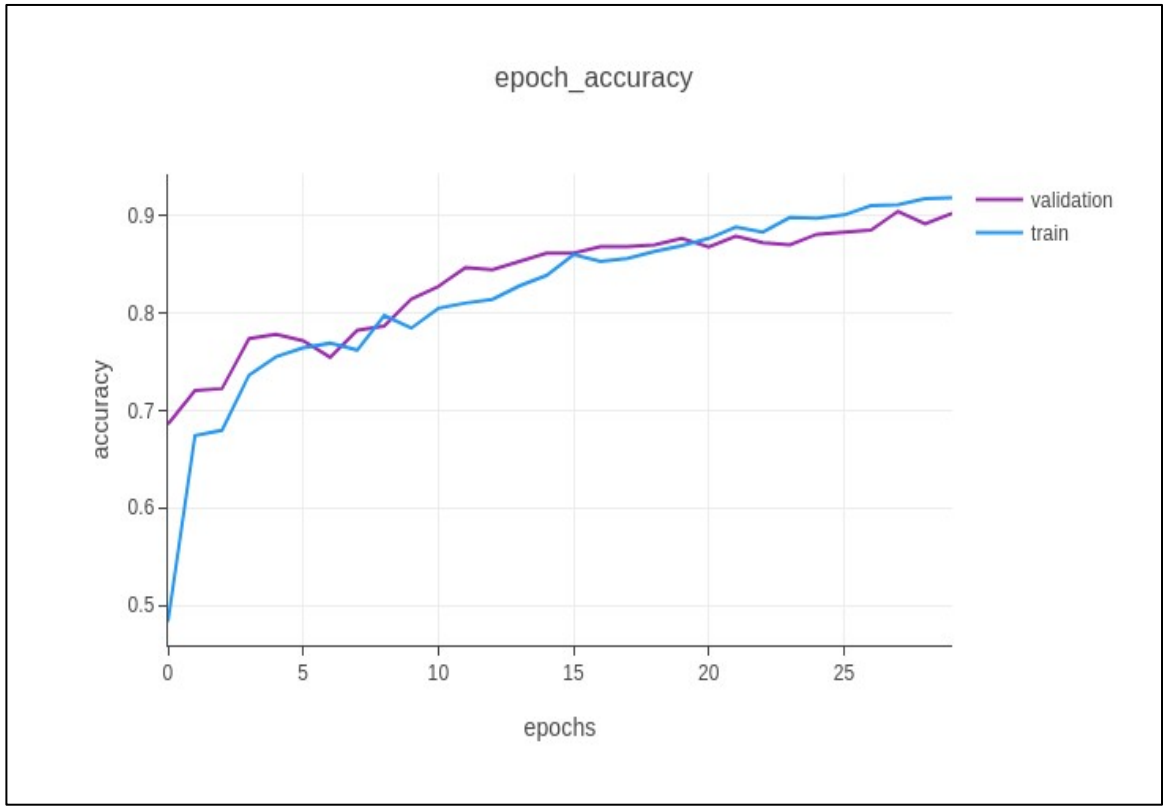
DAM evaluates matching scores for input-answer pairs and returns the response with the highest score to the user.

Model Comparison

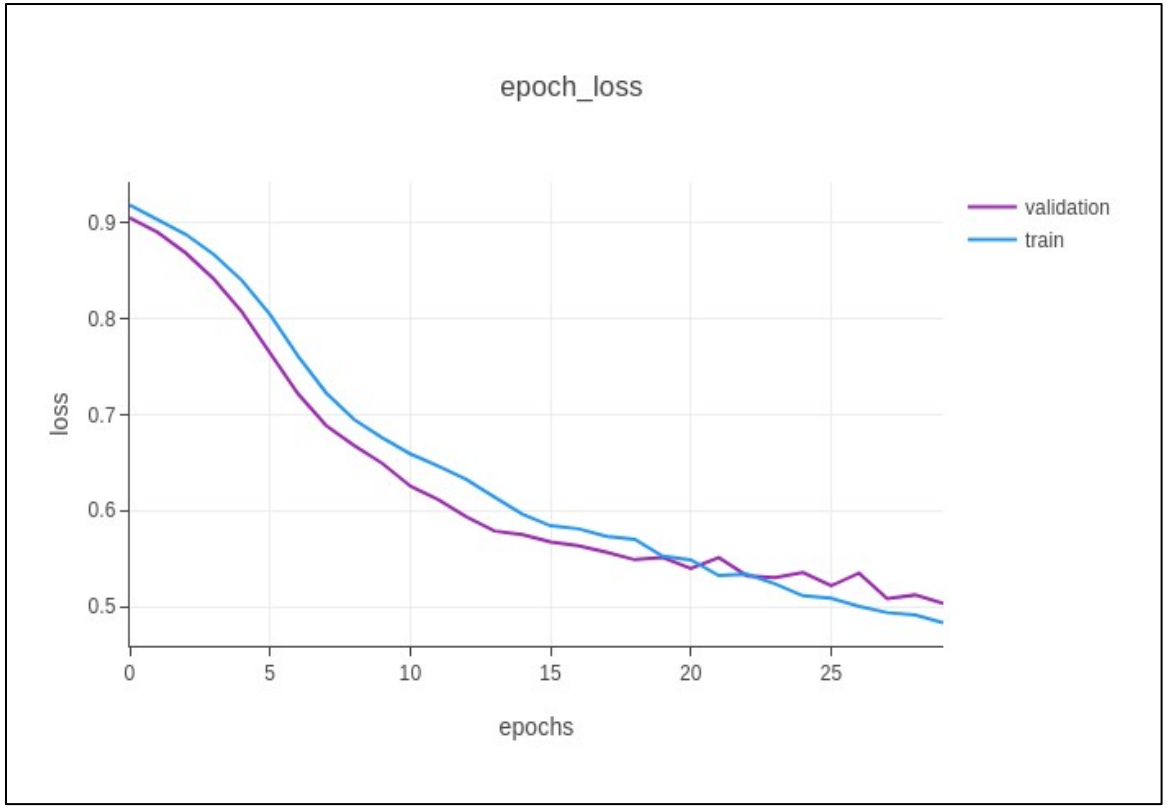
Model	Advantages	Disadvantages
Knowledge Graph	Quick and accurate response	Incapable of managing typographical errors and undefined queries.
BiLSTM	Locate the answers corresponding to the respective questions. Able to handle typographical errors	The accurate answer is present among the options but may not be chosen as the top selection.
DAM	Compute self-attention within a single statement and cross-attention across multiple statements and the response. Choose the optimal response based on attention.	Ineffectiveness in searching for candidate answers.

Siamese Bi-LSTM Model

Accuracy Curves for Bi-LSTM Model

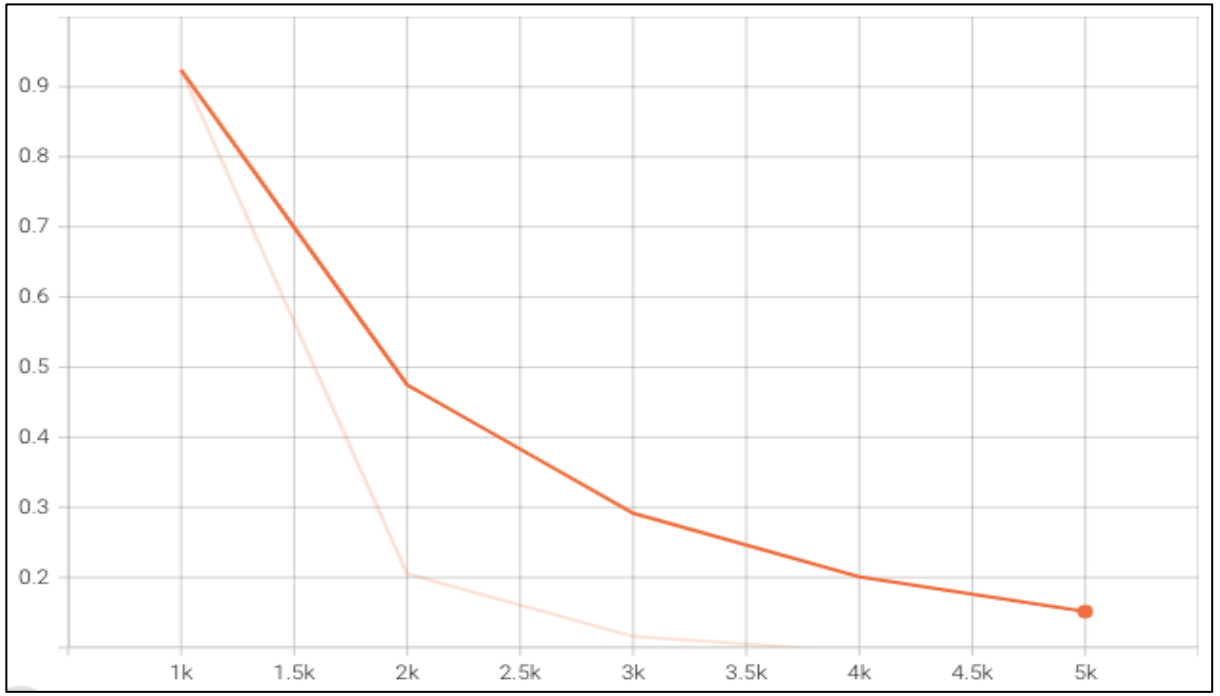


Loss Curves for Bi-LSTM Model



DAM Model

Loss Curve for DAM Model



Evaluation Score for DAM Model

Metrics	Score
Perplexity	113.85
BLEU	92.50
Precision	98.61
Recall	98.68
F1 Score	98.65
Rouge	93.46

YOLO Model

Yolo v5 Evaluation Results

Metric	Values
Number of Epochs	710
Precision	0.744
Recall	0.906
mAP50	0.925
mAP50-95	0.563
Box_loss	0.03454
Obj_loss	0.03918
Class_loss	0.0084

Code implementation for image input

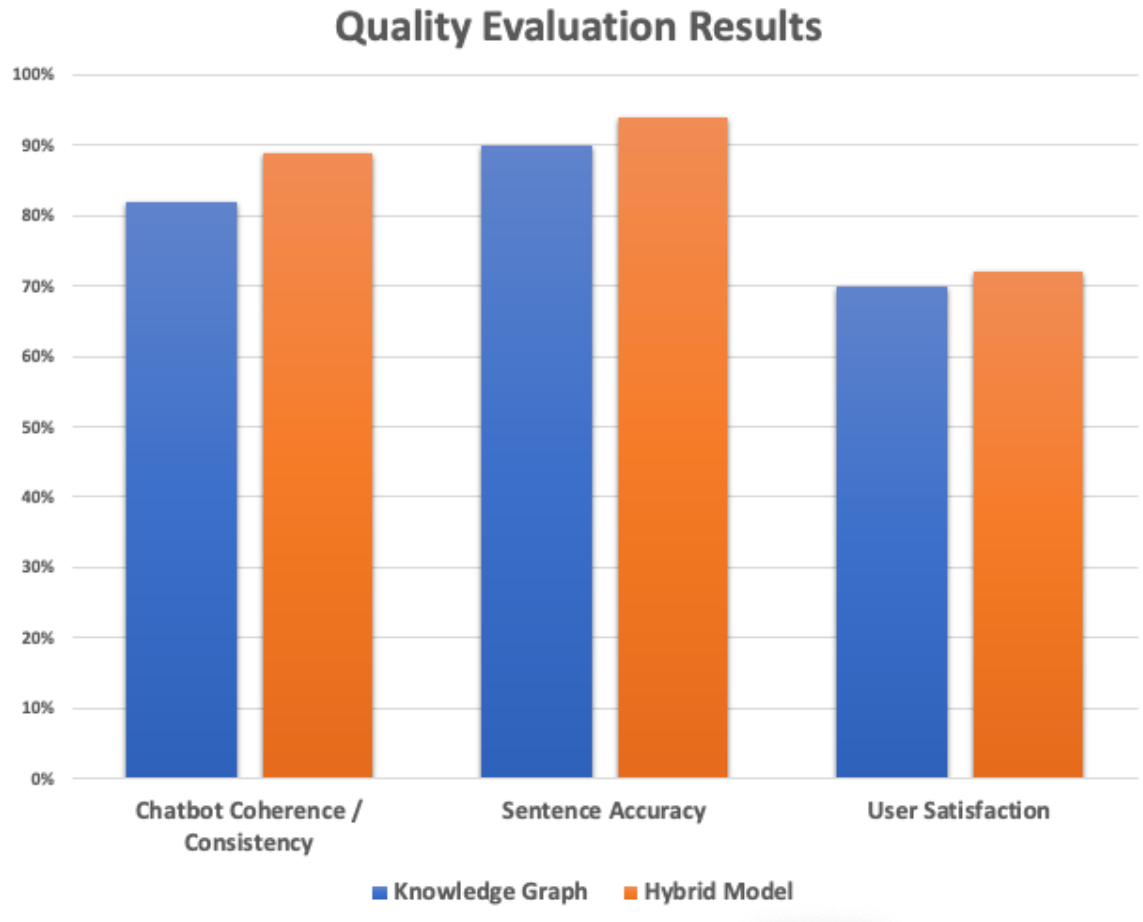
Demo showing GUI element Detection using YOLO and Text Extraction using Tesseract OCR:
https://drive.google.com/file/d/1SPIKjKvPL0i2eQUiIr2q_niMmxYZVnC3/view?usp=drive_link

Backend chatbot demo for Image + text input:
https://drive.google.com/file/d/1ZBF1fArNgq7QxsEF2aAeUFD0mcfVmmwU/view?usp=drive_link

Qualitative Comparison

Overall Comparison of Qualitative Metrics

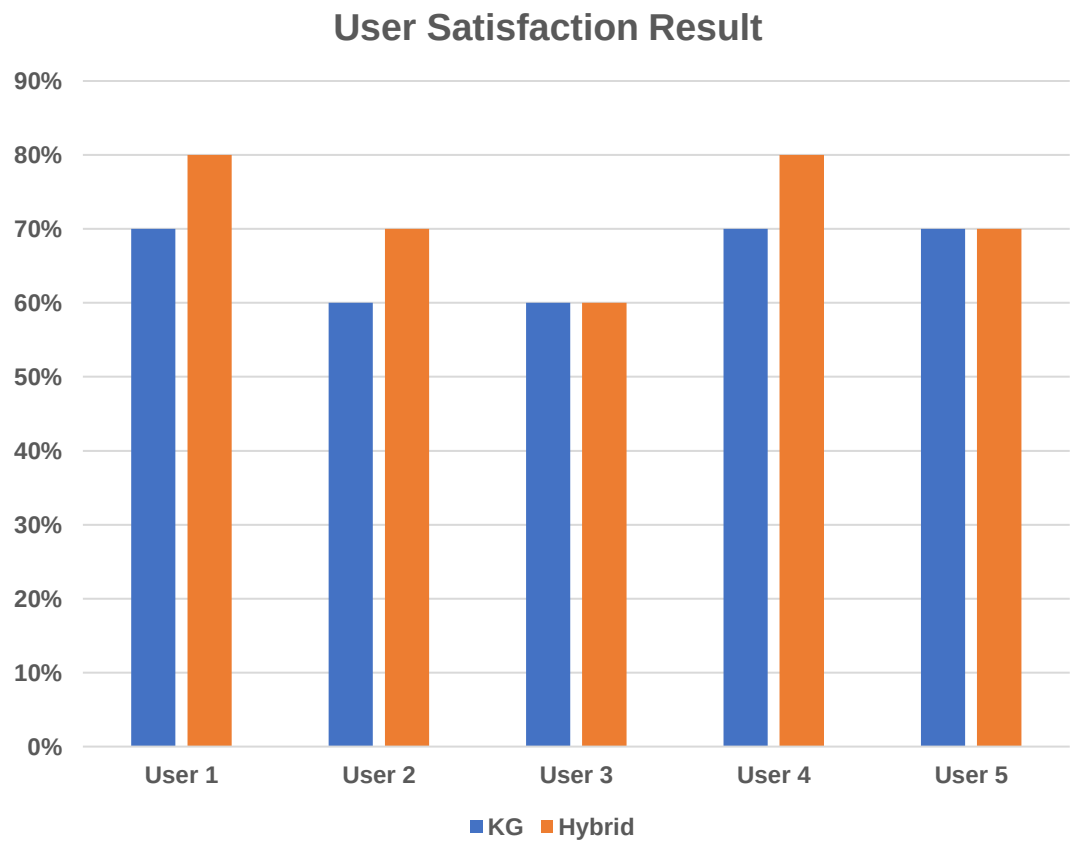
Assessment Metric	Knowledge Graph	Hybrid Model
Chatbot Coherence / Consistency	82%	89%
Sentence Accuracy	90%	94%
User Satisfaction	68%	72%



Qualitative Comparison – User Satisfaction

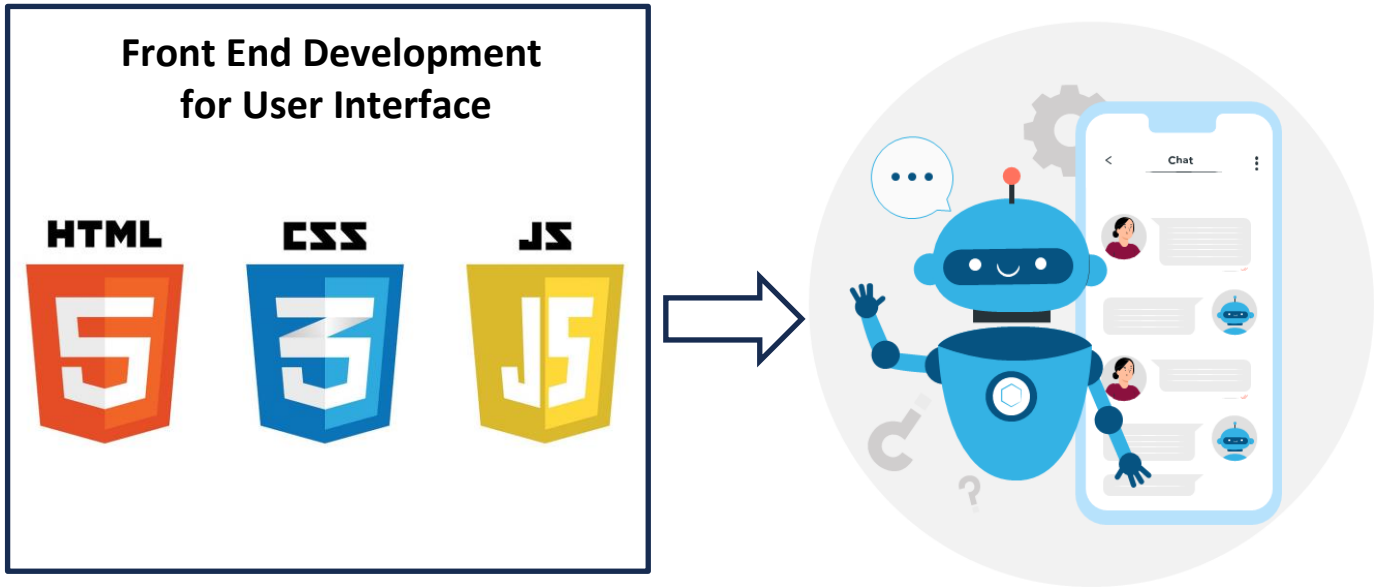
Outcome of the satisfaction Assessment

User	Model	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Satisfaction (%)
1	KG	1	0	1	1	0	1	1	1	1	0	70%
	Hyd	1	1	1	0	1	1	0	1	1	1	80%
2	KG	1	0	0	1	0	1	1	1	0	1	60%
	Hyd	1	0	1	1	0	1	0	1	1	1	70%
3	KG	1	0	1	0	1	1	0	0	1	1	60%
	Hyd	0	1	1	0	1	1	0	1	1	0	60%
4	KG	1	0	1	1	0	0	1	1	1	1	70%
	Hyd	1	1	0	1	0	1	1	1	1	1	80%
5	KG	0	1	1	1	1	0	1	1	0	1	70%
	Hyd	1	0	1	1	0	1	1	0	1	1	70%
Total	**KG**	**4**	**1**	**4**	**4**	**2**	**3**	**4**	**4**	**4**	**4**	**68%**
	Hyd	**4**	**3**	**4**	**3**	**2**	**5**	**2**	**4**	**5**	**4**	**72%**



Web System Design and Development

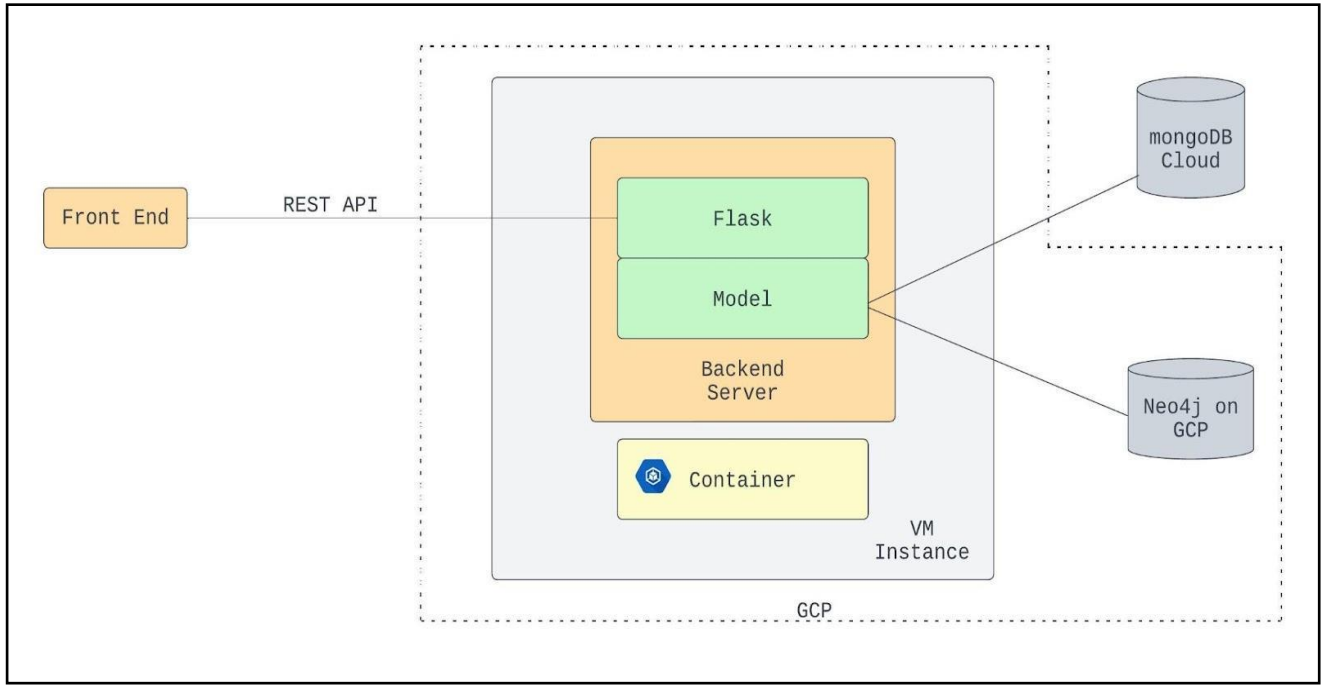
System Infrastructure – User Interface Design



- **HTML**: Constructs the core structure and layout of the user interface for the AI testing tool.
- **CSS**: Applies styling and design elements to the user interface appearance.
- **JavaScript**: Enhances user interface functionalities with advanced features.

System Infrastructure – System Deployment Architecture

- **Cloud Backend**: Hosted on the cloud to provide centralized services.
- **REST API**: Offered by the backend via a Python Flask framework for frontend communication.
- **Chatbot Integration**: AI chatbot embedded within the backend, utilizing Neo4J and MongoDB databases.
- **Containerization**: Backend operates within a container on a GCP virtual machine instance.
- **Database Deployment**: Neo4j database runs on GCP; MongoDB is utilized as a cloud-managed service.



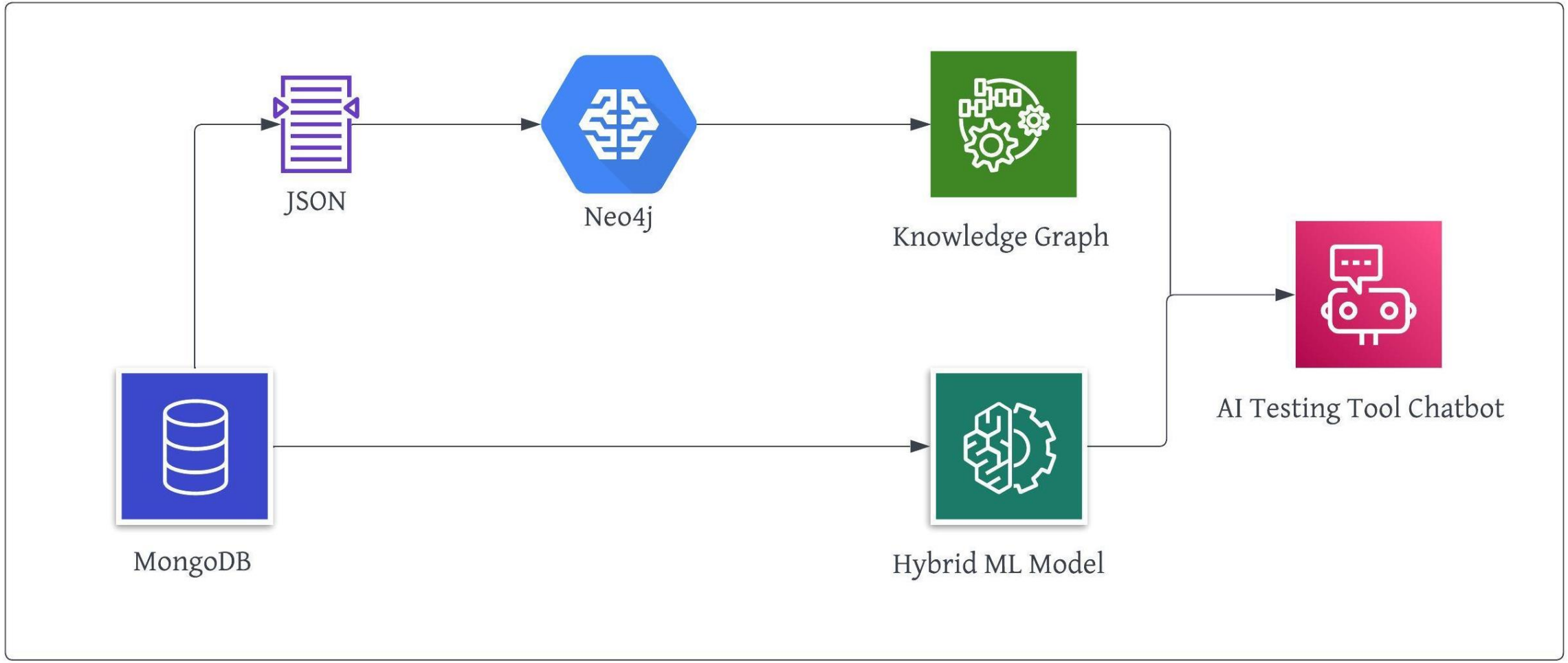
Database System Design

- MongoDB serves as the central repository for both unprocessed and processed data storage.

- The refined data, extracted in JSON format from MongoDB, is channeled into a knowledge graph model.

- Subsequently, the data is fed into a Neo4J graph database, and through an integrated model, the chatbot API retrieves responses.

Database System Design



System Supporting Environment

Components	Environments	Ports
Frontend	Local Web App	5000
Backend Server	GCP VM Instance	3389 -> 89
BiLSTM	Separate GCP VM Instances with different ports mapping to individual containers	3389 -> 80
DAM		7000 -> 22370
Knowledge Graph		5000 -> 5000
Neo4j	GCP VM - Self-Hosted Server	7474
MongoDB	Sourced by Mongo	NA

System Implementation

System Implementation

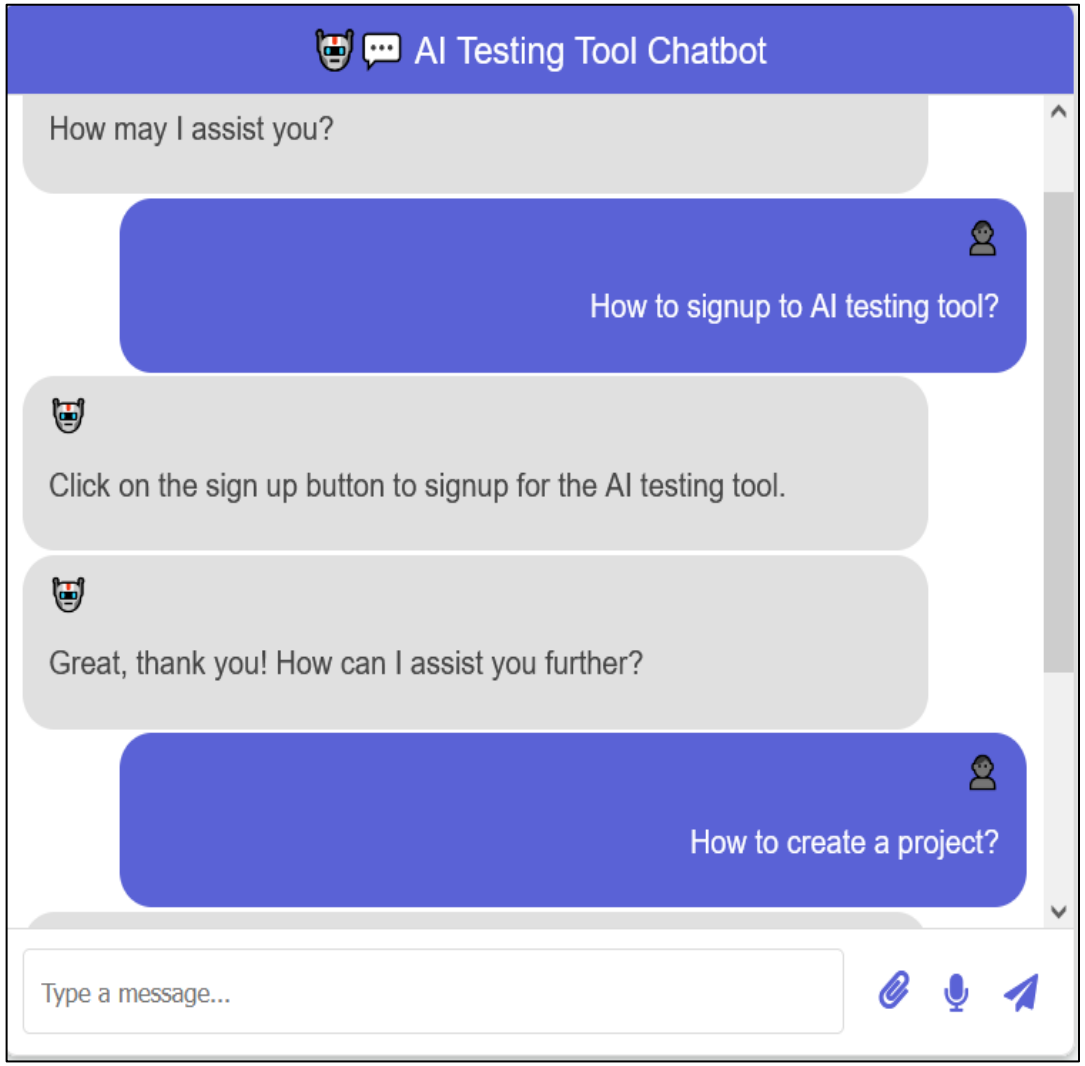
- Python Flask Framework is used to implement web application
- Q&A data is feed to Knowledge Graph and DAM model
- After training the model, relevant response from chatbot is displayed in UI
- System Implementation results include:

    Chatbot Interface for text data, audio and image data

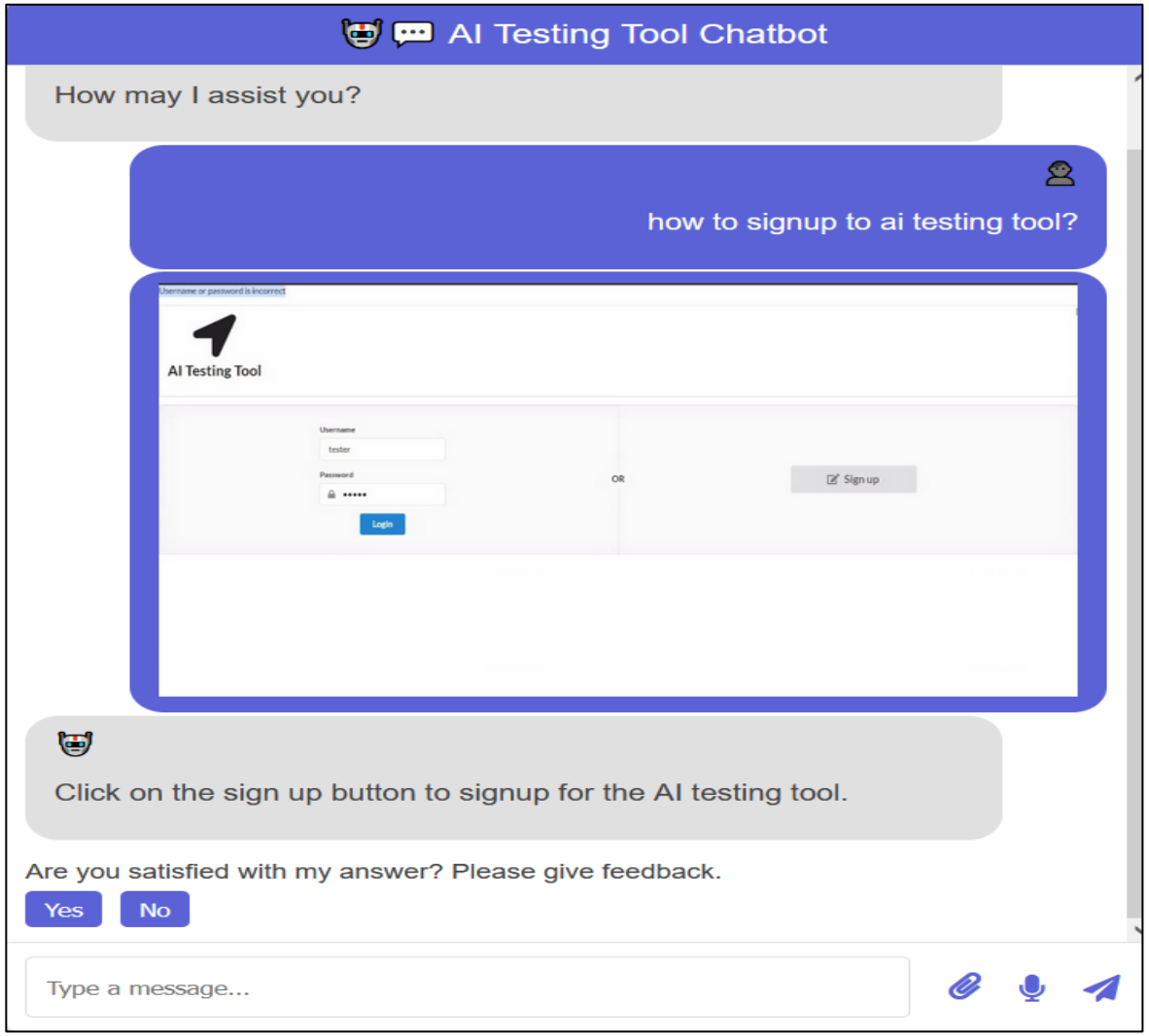
    Application Scenarios: Consistency, Correctness handling and Sentence Accuracy

Chatbot Interface

For Text Data



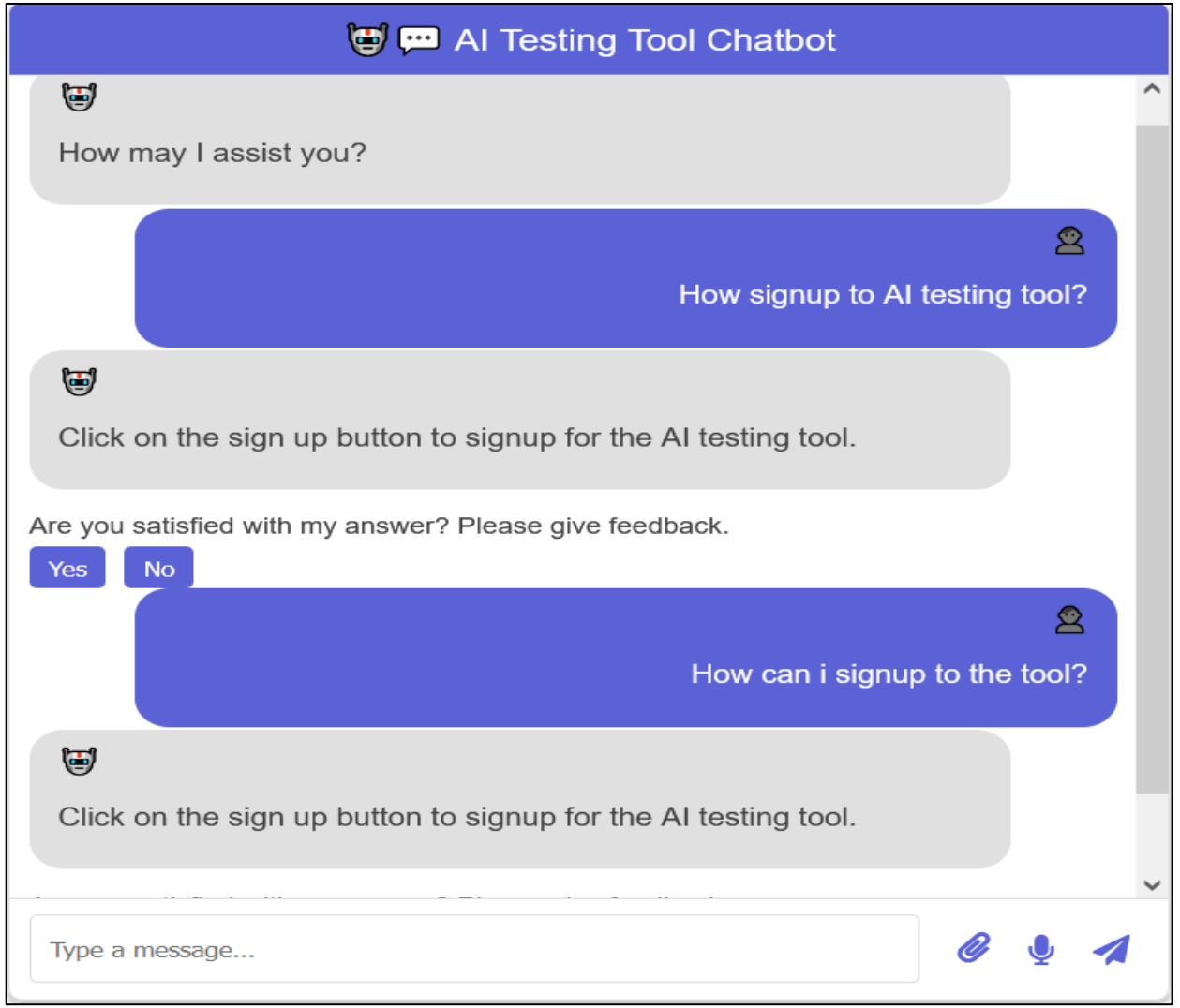
Uploading AI Testing Tool Screenshot



Application Scenarios

Chatbot Consistency

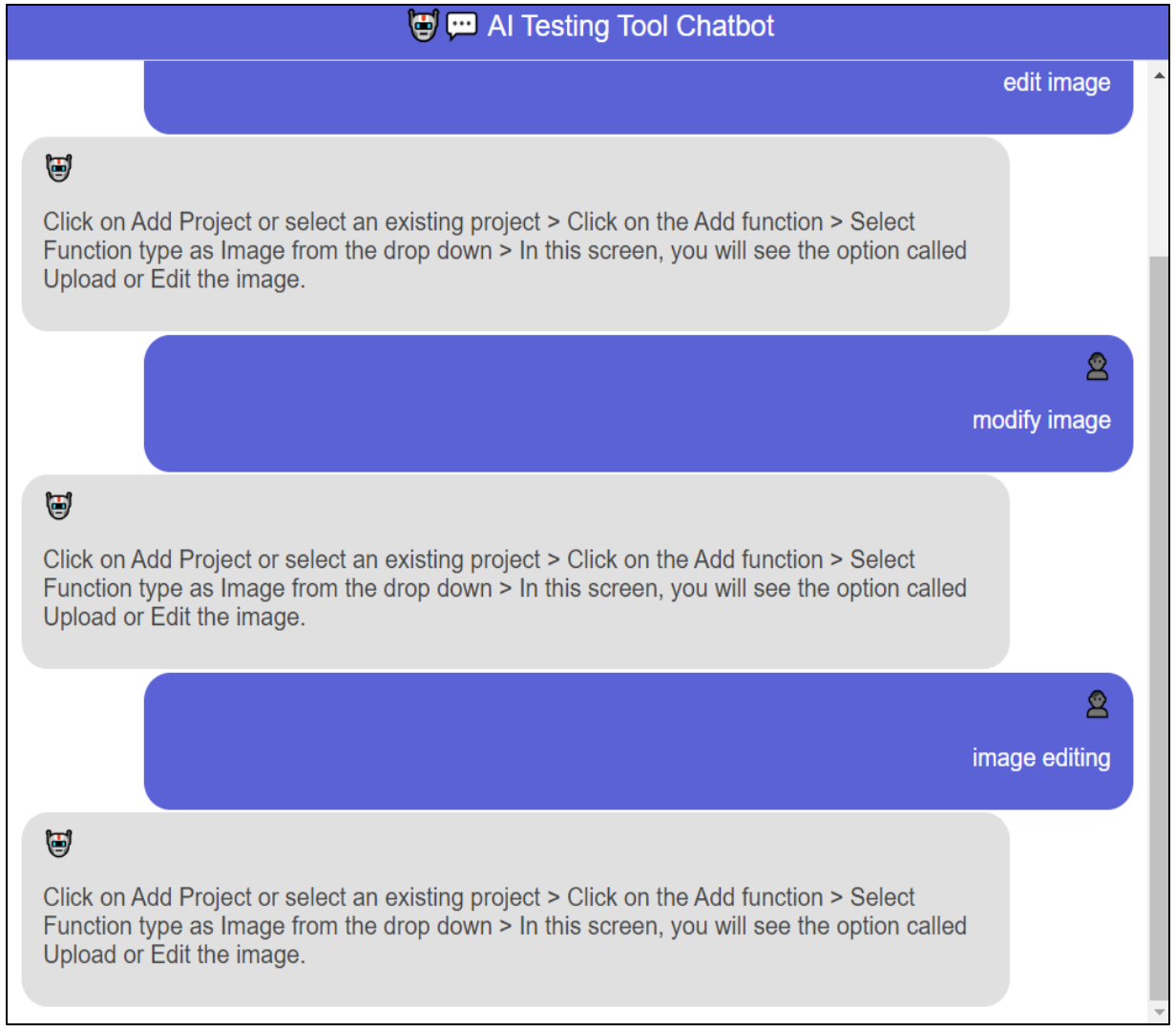
- Testing Chatbot with similar type of questions and getting the same response



Application Scenarios

Sentence Accuracy

- Testing chatbot with spelling and grammatical errors and getting the same response



Demo

Final UI Demo

Demo Link:

https://drive.google.com/file/d/1mgGgMLbC5wmsKEU0knvB76iQbIt6Rpw0/view?usp=drive_link

Future Works

- **Automated Data Collection:** Develop a system for the chatbot to autonomously collect high-quality data during deployment, enhancing its learning capabilities
- **Self-Evaluation Feature:** Integrate a self-assessment mechanism using metrics like BLEU Score, Perplexity, and engagement factors to enable continuous improvement in the chatbot's performance
- **Iterative Learning Loop:** Establish a loop for reusing training data, continuously evolving the core model for improved response accuracy
- **Zoom Integration:** Plan to add a Zoom screen sharing feature for a more interactive user experience in future updates, aiming to boost engagement and satisfaction

Thank you