IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 1, January 2025

DOI: 10.17148/IJARCCE.2025.14148

Portable AI Voice Assistant using Large Language Model, Speech-to-Text and Text-to-Speech

Mrs. Bhagya[1], Rishabh S Mallir[2], Sindhu S[3], Uday Kiran N C[4], Naveen Shankar Devadiga[5]

Professor, Department of Electronics and Communication, East West Institute of technology, Bangalore, India[1]

Student, Department of Electronics and Communication, East West Institute of technology, Bangalore, India[2]

Student, Department of Electronics and Communication, East West Institute of technology, Bangalore, India[3]

Student, Department of Electronics and Communication, East West Institute of technology, Bangalore, India[4]

Student, Department of Electronics and Communication, East West Institute of technology, Bangalore, India[5]

Abstract: This project introduces an intelligent system that integrates custom-trained large language models (LLMs), RFID-based mode switching, and cloud-based APIs to enable natural, context-aware human-machine interaction on resource-constrained devices. The system operates in three modes—Student Mode, General Mode, and Visitors Mode—each tailored to specific user needs, such as educational support, everyday tasks, and quick information retrieval. RFID technology allows seamless mode switching, while cloud APIs handle resource-intensive tasks like speech-to-text (STT) and text-to-speech (TTS), ensuring real-time responsiveness on low-power hardware like microprocessors. Applications span education, IoT, healthcare, and customer support, enabling voice-activated smart devices, interactive kiosks, and accessibility tools. By combining affordability, scalability, and advanced AI capabilities, this project bridges the gap between cutting-edge technology and practical, real-world solutions, making AI-driven systems more accessible and impactful across industries.

Keywords: LLM, RFID, cloud APIs, STT, TTS, IoT, education, healthcare, accessibility, low-power hardware.

I. INTRODUCTION

This project combines **RFID-based AI model switching, voice interaction, and document AI processing** to create an affordable and efficient AI system. It allows users to **automatically switch AI models** using RFID, making tasks easier without manual intervention—like a doctor switching between diagnosis and patient history models. The system also supports **voice-based interaction** using Google Cloud APIs, enabling hands-free communication, which is especially useful for people with disabilities or in fast-paced environments. Additionally, it can **analyse and process documents** using Google Document AI, helping professionals extract important information quickly. This technology can be used in **healthcare, smart homes, education, business, and public services** to improve efficiency and accessibility. By relying on **low-cost hardware and cloud-based processing**, the system remains both **affordable and scalable**, making advanced AI solutions more accessible to individuals, small businesses, and large organizations.

1.1 MOTIVATION

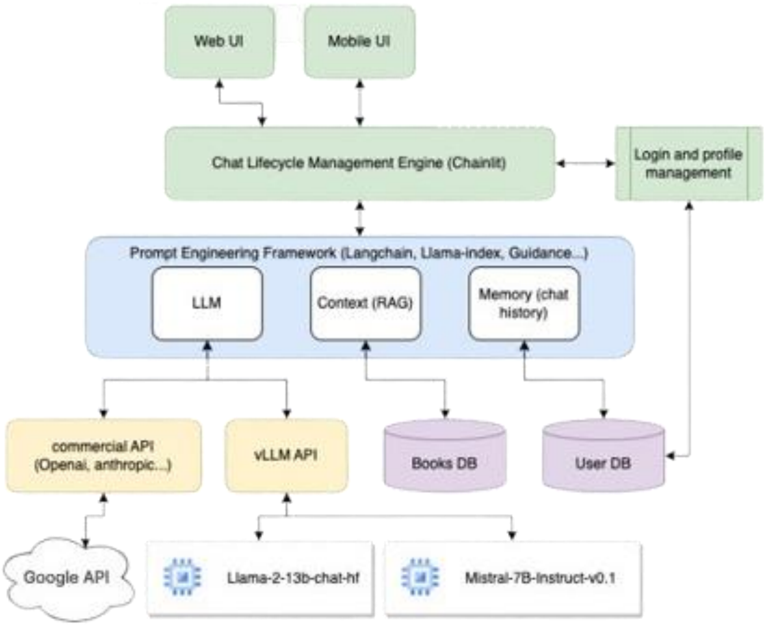
The motivation for this project comes from the growing need for **intelligent, adaptive, and accessible AI systems** that can integrate seamlessly into various industries. Many existing solutions lack **automation, accessibility, and scalability**, making them impractical for small-scale users and resource-limited environments. This project **bridges that gap** by combining **RFID-based AI model switching, voice interaction, and document AI processing** with cloud-based APIs. **RFID technology** enables automatic AI model switching, making tasks like medical diagnosis more efficient. **Voice interaction** using Google Cloud APIs allows hands-free communication, benefiting individuals with disabilities and users in fast-paced settings. **Document AI processing** helps industries like healthcare and education analyze information quickly, reducing manual effort. Since the system **uses low-cost hardware and cloud-based AI**, it remains **affordable and scalable**, making advanced AI accessible to schools, small businesses, and developing regions. Beyond technology, this project also promotes **digital inclusivity** and serves as a learning tool for students exploring **AI, IoT, RFID, and cloud computing**. By enabling **smart, voice-activated interactions on compact devices**, it pushes the boundaries of **IoT and automation**, making AI-driven solutions more practical and widely available.

DOI: 10.17148/IJARCCE.2025.14148

1.2 OBJECTIVE

- Convert LLM text responses into speech for real-time audio output.
- Optimize data transmission for efficient STT processing using cloud APIs.
- Develop a communication pipeline for sending text prompts to an LLM and receiving responses.
- Create a compact, portable device for real-time voice command processing.
- Enable hands-free, voice-based interaction for accessibility and usability.
- Optimize the system for low-cost hardware and resource-constrained environments.
- Train the LLM on domain-specific documents for improved accuracy.
- Ensure scalability and flexibility for diverse applications and deployments.

II. METHODOLOGY


Web UI
Mobile UI
Chat Lifecycle Management Engine (Chainlit)
Login and profile management
Prompt Engineering Framework (Langchain, Llama-index, Guidance...)
LLM
Context (RAG)
Memory (chat history)
commercial API (Openai, anthropic...)
vLLM API
Books DB
User DB
Google API
Llama-2-13b-chat-hf
Mistral-7B-Instruct-v0.1

The system is divided into four core components:

2.1 RFID-Based AI Model Switching

Purpose:

- Automate model selection based on user identity or the type of task being performed.
- Minimize manual intervention when switching between different AI models.

Working Process:

1. **User scans an RFID tag** → The RFID reader captures the unique tag ID.
2. **System retrieves model mapping** → The backend database contains predefined mappings of RFID tags to AI models.
3. **AI model is dynamically updated** → The system loads the corresponding AI model.
4. **Confirmation feedback is provided** → The user is notified (via text/voice) that the model has changed.

Outcome:

- Each user or task can be assigned a **dedicated AI model**.
- The system **adapts in real-time** to different workflows or data sets.
- Ensures a **personalized AI experience** without manual configuration.

DOI: 10.17148/IJARCCE.2025.14148

2.2 Voice-Based Input and Output

Purpose:

- Provide hands-free AI interaction for accessibility and efficiency.
- Enable users to communicate naturally with the AI.

Working Process:

1. **User speaks a command/query** → The system records the speech input.
2. **Speech is converted to text** → Google's Speech-to-Text API processes the input.
3. **AI processes the query** → The active AI model interprets the text and generates a response.
4. **Response is converted back to speech** → The system uses Google's Text-to-Speech API.
5. **User hears the AI-generated voice response** → The system plays the output.

Outcome:

- Users can interact with AI **without typing or clicking**.
- The system can provide **audible responses** in real-time.
- Enhances **accessibility** for users with mobility or visual impairments.

2.3 Document-Specific AI Training

Purpose:

- Improve AI's ability to **analyze and process specific types of documents**.
- Enable the AI model to **extract relevant information** from structured and unstructured data.

Working Process:

1. **User uploads a document** → The system accepts various formats (PDF, DOCX, scanned images).
2. **Document AI extracts key data** → The system preprocesses the document (OCR, NLP-based analysis).
3. **AI model is trained on document structure** → Enhances contextual understanding.
4. **User queries AI about document content** → AI provides answers based on trained data.
5. **System refines results based on feedback** → Continuous learning improves accuracy.

Outcome:

- AI can **understand and extract** structured insights from uploaded documents.
- Enhances **precision in AI-generated responses**.
- Reduces **manual document review effort**.

2.4 Google Cloud API Integration

Purpose:

- Leverage cloud-based AI services for improved speech and document processing.

APIs Used & Their Role:

- **Google Speech-to-Text** → Converts spoken input into text.
- **Google Text-to-Speech** → Converts AI-generated text responses into natural-sounding speech.
- **Google Document AI** → Extracts information from documents and improves AI accuracy.

Working Process:

1. **System detects the required task** → Speech recognition, text analysis, or document processing.
2. **Relevant API is triggered** → Processes the input in real time.
3. **AI model uses API-generated insights** → Enhances decision-making.
4. **System delivers final response** → As text or voice output.

Outcome:

- **Efficient speech processing** for real-time interaction.
- **Enhanced document comprehension** using cloud AI.
- **Scalability** to process large volumes of data without local computation.

DOI: 10.17148/IJARCCE.2025.14148

III. IMPLEMENTATION

3.1 HARDWARE SETUP

3.1.1 Setting up the Voice Assistant Module: The hardware setup begins with the integration of the microprocessor as the primary processing unit. The USB microphone is connected to capture the high quality audio feed. A receiver is attached to facilitate communication with the microprocessor, while a dedicated power supply module ensures consistent power delivery to the Voice Assistant component. The captured audio feed data will later be transmitted to the user interface for processing.

3.1.2 Integrating the USB microphone for audio recognition: The USB microphone module is configured to capture the audio. It recognizes the audio from the user's speech and sends it to a Goggle Speech-to-Text server were the STT server converts the speech to text format.

3.1.3 Connection of Bluetooth speaker: The Bluetooth speaker is connected to the AI model by the Bluetooth connection for reading of the answer that the Gemini model gets as an output for the question being asked.

3.2 SOFTWARE DEVELOPMENT

3.2.1 Designing the interface: The complete custom designing of the interface can be done using the HTML and CSS code formats. The interface can be designed in such a way that the whole settings can be handled by the admin user or the owner for the interface model

3.2.2 Deploying the Local Host: For developing the Local host the API's can be used to integrate the local host with the other advanced service applications for the web interface.

3.2.3 Loading the models: Here few model for the local host can be trained based on our own data also and can be made on the linking of the cloud APIs. The different modes used is the General Mode and the Visitor mode.

- Where it specifies the General mode as the usage model for getting the answers fro all the basic general questions.
- **The Visitor mode is a trained set of data** for availing the data about the college and the visitors can get the easy navigation to the classrooms and all the offices in the college building and it premises related to the college.

3.2.4 Connecting to the cloud: The local host has to be connected to the cloud server for all web searching and resolving the query. The local host can be connected with the cloud using the API keys so that they can establish the better connection between the host and the service application.

3.2.5 Conversion of Text-to-Speech and Speech-to-Text: As the input is given in the audio format and for the web searching it has to be converted into the text format, and the output has to be spelt out in the audio format. So that the Eleven Labs APIs are being used for all the inner conversions.

IV. RESULT

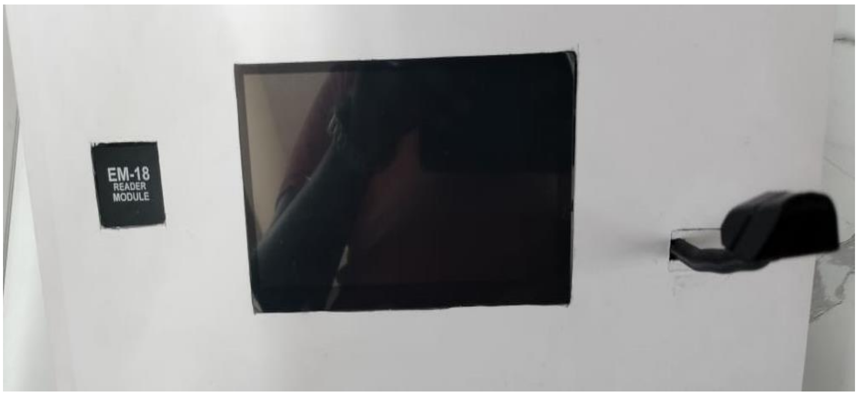
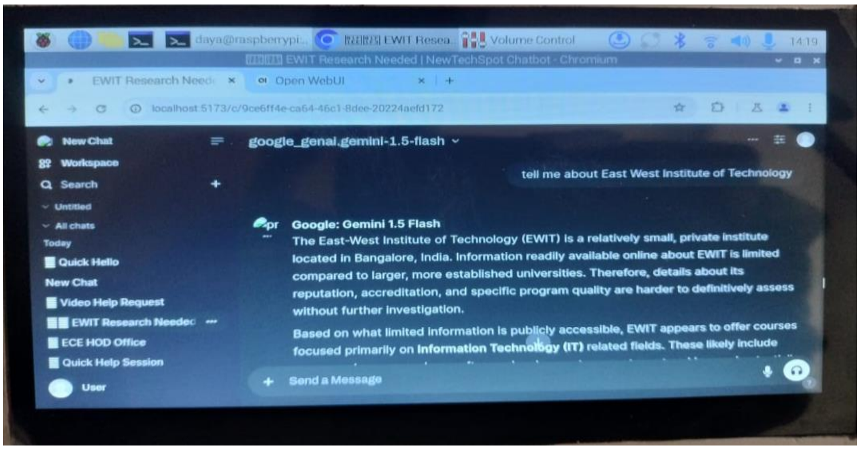


Fig 2. Model of the AI Voice assistant

DOI: 10.17148/IJARCCE.2025.14148



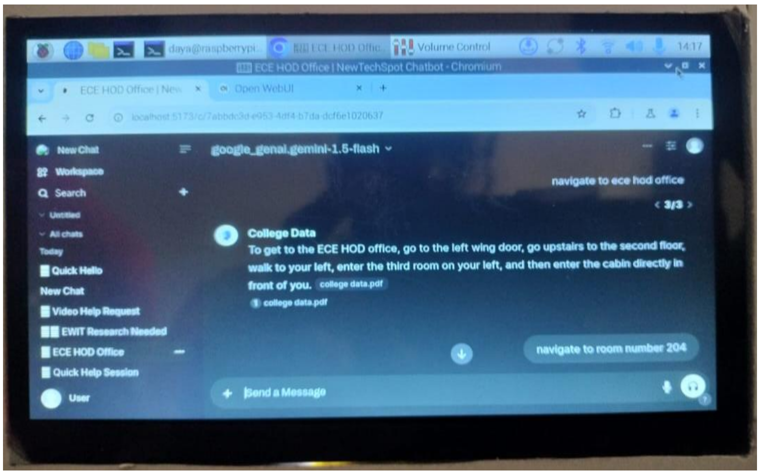
tell me about East West Institute of Technology

Google: Gemini 1.5 Flash

The East-West Institute of Technology (EWIT) is a relatively small, private institute located in Bangalore, India. Information readily available online about EWIT is limited compared to larger, more established universities. Therefore, details about its reputation, accreditation, and specific program quality are harder to definitively assess without further investigation.

Based on what limited information is publicly accessible, EWIT appears to offer courses focused primarily on Information Technology (IT) related fields. These likely include

Fig 3. The output in the General model about general queries



navigate to ece hod office

College Data

To get to the ECE HOD office, go to the left wing door, go upstairs to the second floor, walk to your left, enter the third room on your left, and then enter the cabin directly in front of you. college data.pdf

1 college data.pdf

navigate to room number 204

Fig 4. The output in the visitors model for college information

CONCLUSION

The development of an AI voice assistant using **Raspberry Pi** demonstrates the feasibility of implementing advanced speech-based interaction on a compact, low-power computing platform. By integrating **speech-to-text (STT), text-to-text processing (NLP), and text-to-speech (TTS)** functionalities, the project showcases the potential of Raspberry Pi in real-time voice-based automation.

This project not only validates the capability of Raspberry Pi in handling AI-driven voice processing but also emphasizes its scalability for various applications, such as **smart home automation, hands-free assistance, and personalized AI interfaces**. The successful implementation highlights the efficiency of open-source libraries and AI models in enabling human-like interactions.

With future enhancements, such as **improved speech recognition accuracy, multilingual support, and IoT integration**, this project can serve as a foundation for more sophisticated AI-driven systems. Ultimately, it contributes to the growing field of embedded AI, making voice-enabled computing more accessible and efficient.

FUTURE SCOPE

 Enhanced AI Model Switching – Implement **multi-factor authentication** (RFID + biometrics) and **predictive AI model selection** based on user behavior.

DOI: 10.17148/IJARCCE.2025.14148

□ **Advanced Voice Interaction** – Add **multilingual support**, **context-aware conversation AI**, and **background noise filtering** for better voice recognition.

□ **Improved Document Processing** – Enable **automated document summarization**, **handwritten text recognition**, and **real-time data extraction**.

□ **Integration with IoT & Smart Systems** – Apply the system in **smart homes, hospitals, industrial automation**, and **wearable AI devices**.

□ **Industrial & Commercial Applications** – Expand to **healthcare (RFID-based patient ID)**, **retail (inventory automation)**, and **education (personalized AI tutors)**.

□ **Security & Privacy Enhancements** – Implement **end-to-end encryption**, **AI bias reduction**, and **blockchain-based AI auditing** for secure data handling.

□ **Hybrid Cloud & Edge Computing** – Deploy AI models on **edge devices (Raspberry Pi, Jetson Nano)** to reduce cloud dependency and improve processing speed.

□ **Real-Time AI Adaptation** – Improve **AI learning mechanisms** so that models dynamically adapt to user **preferences, tasks, and document content**.

□ **Scalability & Custom AI Trasining** – Automate **AI model retraining** using continuous **document analysis and machine learning pipelines**.

□ **Blockchain for AI Security & Logging** – Use blockchain to **secure RFID transactions, track AI model switching**, and ensure **tamper-proof AI decision records**.

REFERENCES

1.A Survey on LLM-Generated Text Detection:Necessity, Methods, and Future Directions(2024) ,Junchao Wu,Runzhe Zhan, Shu Yang,Derek Fai Wong∗,Lidia Sam Chao NLP2CT Lab, Faculty of Science and Technology.
2. From Summary to Action: Enhancing Large Language Models for ComplexTasks with Open World APIs published in feb-2024 Yulong Liu1,Yunlong Yuan2,Chunwei Wang3 Jianhua Han3 Yongqiang Ma1Li Zhang2 Nanning Zheng1, Hang Xu(Referred byarXiv:2402.18157v1)
3. OPEN-SOURCE LLMS FOR TEXT ANNOTATION: A PRACTICAL GUIDE FOR MODEL SETTING AND FINE-TUNING(2024) by Meysam Alizadeh ,MaëlKubli,ZeynabSamei,Shirin Dehghani,MohammadmasihaZahedivafa,Juan D. Bermeo,Maria Korobeynikova Fabrizio Gilardi.
4. A Survey on Large Language Models: Applications, Challenges,Limitations, and Practical Usage.published in October 2023 by Muhammad Usman Hadi ,qasem al tashi , Rizwan Qureshi , Abbas Shah , amgadmuneer , Muhammad Irfan , Anas Zafar , Muhammad Bilal Shaikh , Naveed Akhtar ,Jia Wu , and SeyedaliMirjalili(Referred byarXiv).