International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Multilingual AI Assistant

Shreya Dewangan[1], *Sonali Sahu*[1], *Samta Gajbhiye*[2]

[1] Student, Department of Computer Science, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh, India 490006
[2] Assistant Professor, Department of Computer Science, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh, India 490006

ABSTRACT:

Language barriers often hinder effective communication in our increasingly globalized world. This project introduces a Multilingual AI Assistant designed to understand and respond to user inputs across multiple languages. Leveraging speech-to-text (STT) conversion, a Large Language Model (LLM), and text-to-speech (TTS) synthesis, the assistant provides real-time, context-aware interactions. Developed using Python and integrated with the Google Gemini API, the system offers both textual and auditory responses, enhancing user engagement and accessibility. This paper presents the conceptual framework, system architecture, implementation strategy, evaluation metrics, and potential real-world applications of the assistant in fields such as education, healthcare, and customer service.

Keywords: Artificial Intelligence, Multilingual Assistant, Speech Recognition, Text-to-Speech, Natural Language Processing, Large Language Models

Introduction

In the modern era of globalization and digital transformation, seamless communication across different languages has become an essential requirement. However, most existing virtual assistants, such as Siri, Google Assistant, and Alexa, are limited in their multilingual capabilities, often requiring manual language switching or offering poor accuracy in less common languages. This limits their effectiveness in diverse societies where users may switch languages or dialects within a single conversation.

To address these limitations, our project aims to design and develop a Multilingual AI Assistant capable of understanding and interacting with users in multiple languages, both in text and speech formats. By utilizing recent advancements in natural language processing (NLP), speech recognition, and large language models, the assistant ensures a fluid and intelligent user experience.

The system integrates speech-to-text (STT) and text-to-speech (TTS) technologies to support spoken communication, while transformer-based models like Google Gemini and OpenAI GPT provide deep contextual understanding across languages. Additionally, the assistant leverages open-source libraries and cloud APIs to enable real-time language translation, multilingual response generation, and dynamic code-switching detection.

A key focus of this project is maintaining user privacy and accessibility. Unlike many commercial solutions that rely heavily on cloud infrastructure, our assistant is designed to function with minimal cloud dependency, ensuring secure data handling and the possibility of offline or edge deployment. This makes it suitable for use in regions with limited internet connectivity or strict privacy requirements.

By supporting seamless, intelligent interaction across languages, this assistant has applications in education, customer support, healthcare, and public services, especially in linguistically diverse regions. Ultimately, this project bridges the language gap in human-computer interaction, promoting inclusivity and digital equity.

Problem Definition and Identification

Despite significant technological advances in virtual assistants, language support remains a persistent challenge. Traditional systems often rely on predefined grammar-based models or only support a few global languages, neglecting linguistic diversity.

The core problems identified are:

1. Limited support for regional or less widely spoken languages.
2. Inability to handle dynamic language switching during conversation.
3. Minimal contextual understanding across different languages.

Our proposed solution aims to overcome these challenges by building a robust, real-time AI system with integrated multilingual support for both input and output.

Literature Review

Various researchers and developers have explored the integration of multilingual capabilities into AI systems.
Martins et al. (2020) developed the MAIA project, showcasing a multilingual AI agent for customer support [1]. Pavitra et al. (2020) reviewed voice assistants with
multilingual support, highlighting challenges in speech processing and translation [2]. Kumar et al. (2021) proposed a multilingual voice assistant using AI to improve real-time communication and reduce latency [5].
Recent work by Paul et al. (2023) introduced a two-way voice assistant using transformer models like BERT and GPT for better contextual understanding [3]. Ahmed (2023) demonstrated the use of Google Gemini with speech recognition and cloud APIs to enhance multilingual interaction, while noting concerns about cloud dependency [4].

Our research builds on these efforts by combining open-source tools, cloud APIs, and privacy-focused techniques to develop a secure and scalable multilingual AI assistant. Recent works have shifted toward incorporating transformer-based language models like BERT, GPT, and Gemini, which offer superior contextual understanding. Additionally, commercial tools like Google Cloud Speech API and Microsoft Azure TTS have enabled more accurate voice-to-text and text-to-speech conversions. However, these tools often require extensive training data or cloud dependency, which may not be suitable for all applications.

Hardware and Software Requirements

Hardware:

- Personal computer with Windows 11 OS
- Minimum 8GB RAM and dual-core processor
- Microphone and speaker for audio I/O

Software:

- Python 3.12.0
- Streamlit (for front end and user interface)
- Google Gemini API (for LLM integration)
- Speech Recognition (Python library for STT)
- pyttsx3 / gTTS (Text-to-Speech engines)
- Python-dotenv (for managing environment variables and securing API keys)
- Additional libraries: NumPy, Requests, LangDetect

Methodology

The assistant is developed using a modular pipeline consisting of the following stages:

1. **Input Layer:**
   - The user initiates the conversation by speaking into the microphone.
   - The input is captured and stored temporarily.
2. **Speech-to-Text Conversion:**
   - The captured audio is processed using the Speech Recognition library or Google Speech API.
   - Language detection is optionally applied using LangDetect to determine the spoken language.
3. **Language Understanding and Processing:**
   - The transcribed text is passed to the Gemini LLM, which processes the query and generates a suitable response.
   - Multilingual capabilities are handled via the model's internal tokenizer and contextual language embeddings.
4. **Response Generation (Text and Audio):**
   - The AI-generated response is converted to speech using pyttsx3 or gTTS.
   - Simultaneously, the response is displayed in textual form on the UI.
5. **User Interface Layer:**
   - Streamlit is used to develop a web-based interface.
   - The interface shows user input, detected language, and response text, and plays the audio response.

Implementation

Implementation was done on a Windows 11 machine with a Python-based environment. The UI was created using Streamlit, allowing for real-time updates and interactivity. Environment variables and API keys were stored securely using the dotenv package. The Gemini API was used to handle natural language queries, leveraging its multilingual capabilities for generating contextual responses.

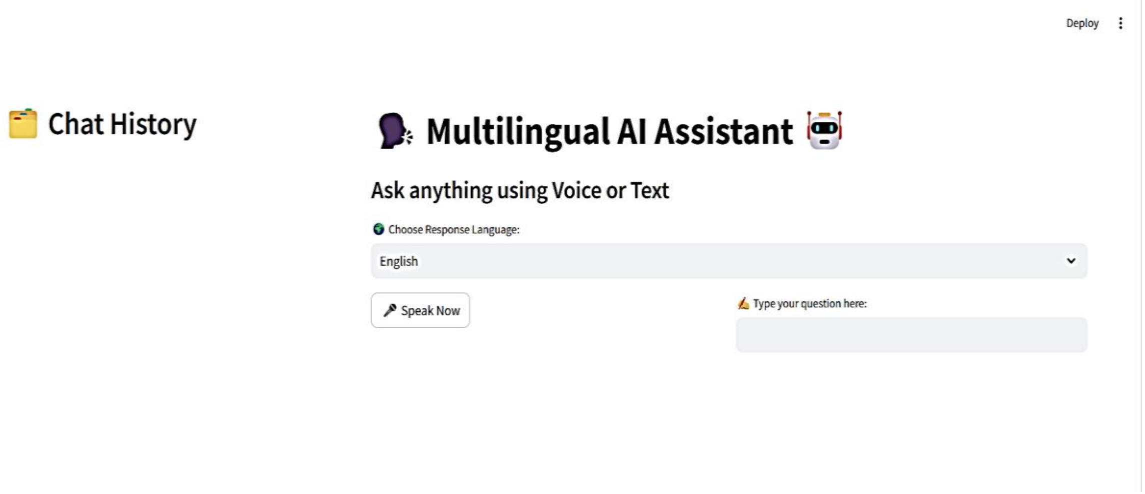

Deploy
Chat History
Multilingual AI Assistant
Ask anything using Voice or Text
Choose Response Language:
English
Speak Now
Type your question here:

Fig1. Landing page.

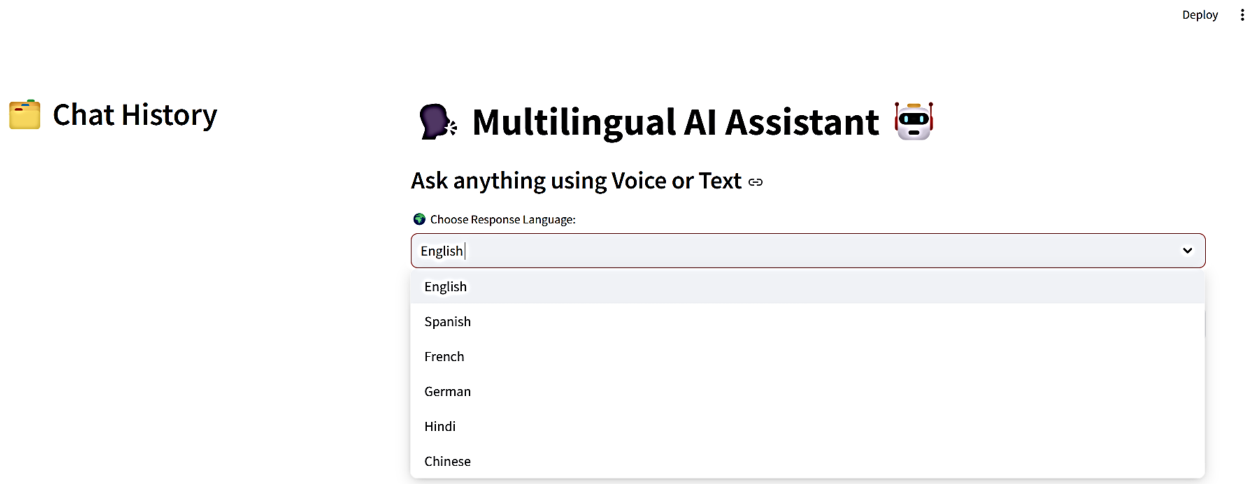

Deploy
Chat History
Multilingual AI Assistant
Ask anything using Voice or Text
Choose Response Language:
English
English
Spanish
French
German
Hindi
Chinese

Fig3. Speech Recognition for input.

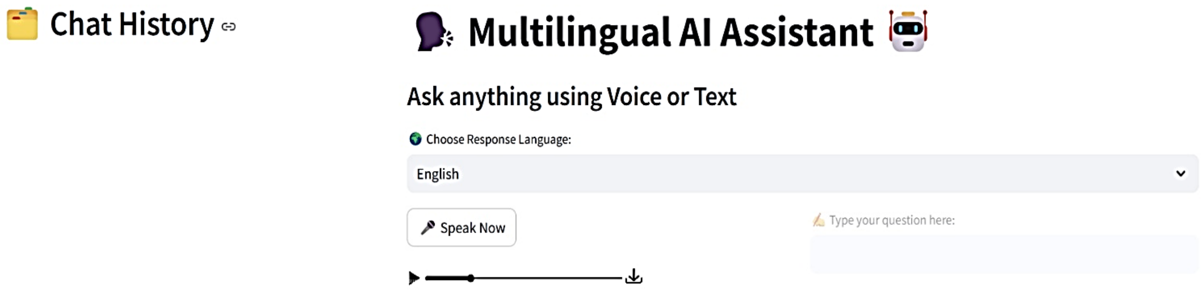
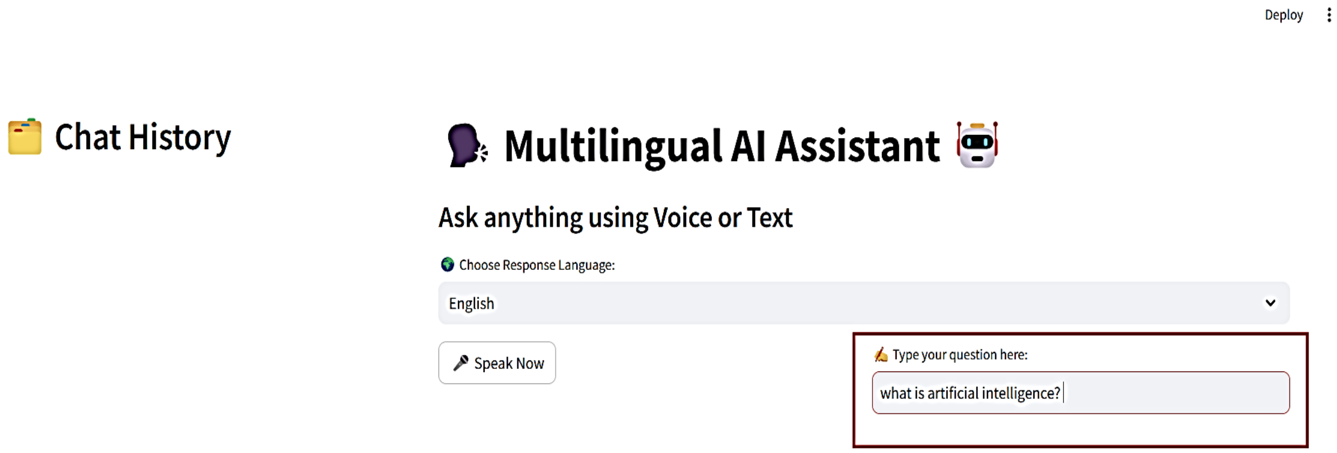

Chat History
Multilingual AI Assistant
Ask anything using Voice or Text
Choose Response Language:
English
Speak Now
Type your question here:

Fig4. AI response in speech.



Chat History

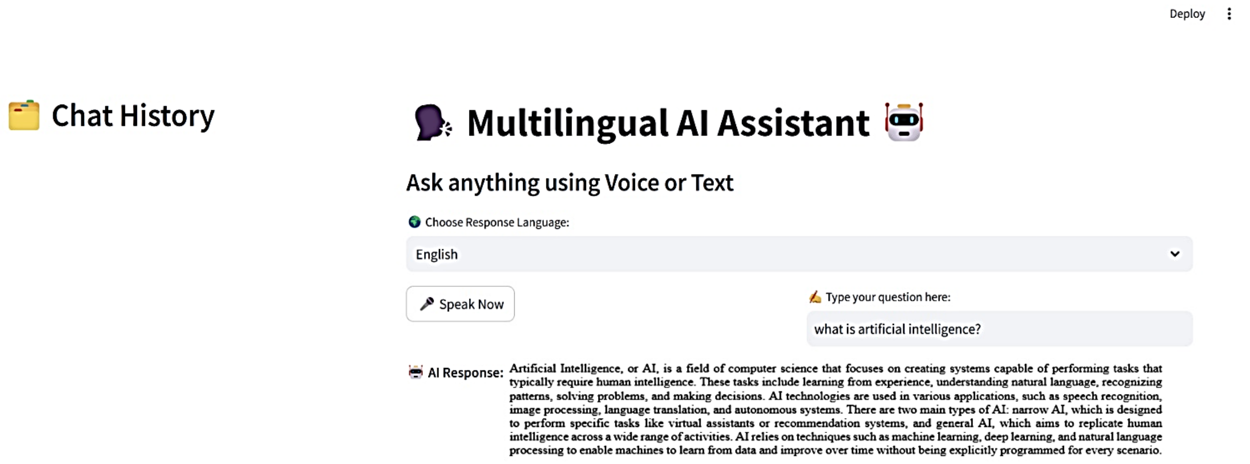
Multilingual AI Assistant

Ask anything using Voice or Text

Choose Response Language: English

Type your question here: what is artificial intelligence?

Fig5. Text Recognition for input.



Chat History

Multilingual AI Assistant

Ask anything using Voice or Text

Choose Response Language: English

Type your question here: what is artificial intelligence?

AI Response: Artificial Intelligence, or AI, is a field of computer science that focuses on creating systems capable of performing tasks that typically require human intelligence. These tasks include learning from experience, understanding natural language, recognizing patterns, solving problems, and making decisions. AI technologies are used in various applications, such as speech recognition, image processing, language translation, and autonomous systems. There are two main types of AI: narrow AI, which is designed to perform specific tasks like virtual assistants or recommendation systems, and general AI, which aims to replicate human intelligence across a wide range of activities. AI relies on techniques such as machine learning, deep learning, and natural language processing to enable machines to learn from data and improve over time without being explicitly programmed for every scenario.

Fig6. AI response in text.

Results and Discussion

The Multilingual AI Assistant was evaluated using inputs in six languages: English, Hindi, Spanish, French, German, and Chinese. The system successfully detected the input language, transcribed the speech accurately, and generated appropriate responses in the selected language. Language detection and switching were handled effectively within the capabilities of the integrated libraries, demonstrating the robustness of the underlying NLP and speech-processing components. The text-to-speech (TTS) output was generally natural, clear, and easily understandable across all tested languages.

However, several challenges were observed during testing. One key issue was handling code-switching within a single conversation, such as users alternating between English and Hindi. While the assistant managed basic switches, maintaining contextual coherence across mixed-language inputs proved difficult. Another limitation was managing API rate limits during continuous or high-volume usage, which occasionally affected real-time responsiveness. Additionally, TTS pronunciation clarity for some Indian languages, especially in region-specific accents, required further fine-tuning to enhance user experience.

Overall, the assistant demonstrated effective multilingual communication, though further optimization is needed for seamless handling of mixed-language input, accent variation, and improved offline functionality.

Conclusion

This research successfully demonstrates the development of a real-time, AI-powered multilingual assistant. The system bridges communication gaps and enhances user experience by enabling fluid interaction in multiple languages. By integrating speech recognition, LLM-based language processing, and text-to-speech synthesis, the project provides a comprehensive solution for multilingual assistance.

Future improvements include:

- Adding support for more Indian and international languages.
- User profiles for personalized experiences.
- Reducing dependency on external APIs by training lightweight local models.
- Improving contextual memory and continuous dialogue tracking.

This work lays the foundation for next-generation AI systems capable of inclusive, accessible, and intelligent communication.

REFERENCES

[1] Martins, A. et al. (2020). Project MAIA: Multilingual AI Agent Assistant. Proceedings of the 22nd Annual Conference of the European Association for Machine Translation.

[2] Pavitra, A. R. et al. (2020). A Review on Intelligent Voice Assistant with Multilingual Support. JETIR, 7(4).

[3] Paul, S. et al. (2023). Two-way Multilingual Voice Assistance. AIJMR, 8(2).

[4] Ahmed, A. (2023). Building Multilingual AI Assistant with Speech Recognition and Google Gemini. LinkedIn.

[5] Kumar, A. K. S. et al. (2021). Artificial Intelligence Based Multilingual Voice Assistant. IJARSCT, 2(3)