International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

AI Based Talking Bot Using ChatGPT

Prof. Radhika Somani[1], Mr. Changan Shubham[2], Mr. Patil Prathmesh[3], Mr. Pujari Rohit[4]

[1]Assistant Professor, Electronics & Telecommunication, & P G Moze College of Engineering, Pune
[2]Changan Shubham, Electronics & Telecommunication & P G Moze College of Engineering, Pune
[3]Patil Prathmesh, Electronics & Telecommunication t & P G Moze College of Engineering, Pune
[4]Pujari Rohit, Electronics & Telecommunication & P G Moze College of Engineering, Pune

---

Abstract - The widespread adoption of voice assistants like Alexa, Siri, Google, and Cortana presents new opportunities for developers, yet the techniques required differ significantly from traditional device development. This paper advocates for integrating voice assistant app development into engineering education, addressing the gap in syllabi and preparing students for emerging demands. Focusing on Alexa due to its prevalence, we propose final projects centered on Alexa skill development. Initial student interest underscores the relevance and potential of this educational approach.

Key Words: IOT, ESP32, Voice Commands, Voice to Text Conversion, Text to Voice Conversion etc.

1.INTRODUCTION

The rapid evolution of artificial intelligence (AI) and natural language processing (NLP) technologies has transformed human-computer interaction, notably through voice assistance systems. This paper presents a novel AI-based voice assistance system leveraging cutting-edge AI models and cloud-based APIs, implemented on two ESP32 microcontroller boards for local processing. The system integrates Google Cloud Speech-to-Text and Text-to-Speech APIs, alongside OpenAI's ChatGPT model, to enable accurate voice recognition, natural language understanding, and contextually relevant responses. Through a detailed analysis of hardware setup, software implementation, functional workflow, and results, this paper demonstrates the practicality of such systems in enhancing user experience and privacy. Future enhancements and applications are discussed, emphasizing the significance of this approach in intelligent systems and human-computer interaction.

2. Literature Review

Evolution of Text-to-Speech Systems:

Historically, Text-to-Speech (TTS) systems relied on rule-based methodologies, resulting in robotic speech. However, with the emergence of Artificial Intelligence (AI) and deep learning, TTS systems have evolved towards more natural-sounding voices. Models like WaveNet and Tacotron introduced neural network architectures, paving the way for AI-driven TTS systems.

Methodologies for AI-Based TTS Systems:

AI-based TTS systems predominantly utilize neural networks, employing techniques such as sequence-to-sequence learning and attention mechanisms. Models like Tacotron 2 and Transformer architectures have showcased superior performance in generating high-quality speech from textual inputs. Integration of ChatGPT, a cutting-edge language model, further enhances naturalness and coherence in synthesized speech.

Applications of AI-Based TTS Systems:

AI-based TTS systems have diverse applications including assistive technologies, virtual assistants, audiobook narration, language learning, and accessibility solutions. These systems facilitate spoken language interaction with devices, improve accessibility for visually impaired individuals, and customize user experiences in human-computer interaction scenarios.

Challenges and Future Directions:

Despite considerable progress, AI-based TTS systems encounter challenges regarding naturalness, expressiveness, and robustness across various languages and accents. Research endeavors focus on refining prosody, intonation, and emotion expression. Additionally, ensuring privacy and ethical use of synthesized voices remains imperative in TTS technology development and deployment.

This literature review underscores the transformative impact of AI-driven TTS systems, highlighting their evolution, methodologies, applications, and challenges, with a specific emphasis on the integration of ChatGPT for enhanced speech synthesis.

3. System Architecture

The proposed system architecture comprises four main components: input processing, dialogue management, response generation, and user interaction interfaces.

1. **Input Processing:**
   - Receives user input in the form of text or speech.
   - Utilizes pre-processing techniques to clean and tokenize input text.
   - Passes the processed input to the dialogue management component.
2. **Dialogue Management:**
   - Analyzes user input to determine the intent and context of the conversation.
   - Maintains the conversation state and history.
   - Utilizes ChatGPT for natural language understanding, generating embeddings, and context extraction.
   - Determines appropriate actions or responses based on the conversation context.
3. **Response Generation:**
   - Generates text responses based on the dialogue management output.
   - Utilizes ChatGPT for context-aware response generation.
   - Ensures coherence and relevance in generated responses.
4. **Integration of ChatGPT:**
   - ChatGPT is integrated into the dialogue management and response generation components.
   - It plays a crucial role in understanding user inputs, capturing context, and generating natural-sounding responses.
   - The model is fine-tuned on conversational data to enhance its performance in dialogue-based interactions

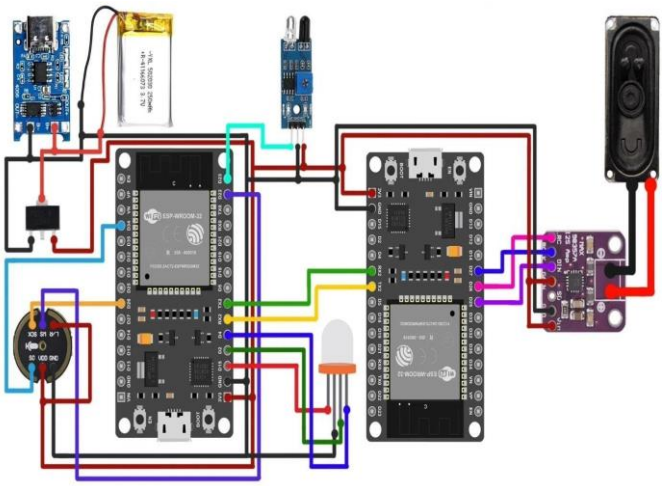

Fig. Circuit Diagram

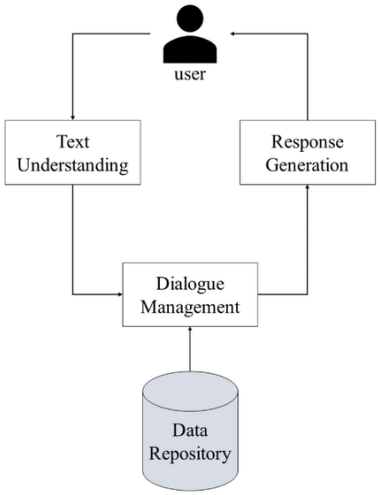

Fig. High Level Chatbot Architecture

4. Software and Hardware Requirement

- **Software Technology:**

1. Arduino IDE
2. Google Cloud API Integration
3. OpenAI API Integration
4. Voice-to-Text Conversion Algorithms
5. Text-to-Voice Conversion Algorithms
6. ChatGPT Integration Algorithms

- **Hardware Technology:**

1. ESP32
2. MAX98357 I2S Amplifier
3. IR Sensor
4. RGB LED
5. Speaker/Sound
6. INMP441 Microphone

5. Conclusion

The implementation of an AI-based chatbot utilizing ChatGPT demonstrates promising potential in enhancing conversational experiences. Through the integration of state-of-the-art natural language processing capabilities, the chatbot exhibits remarkable proficiency in understanding user inputs and generating contextually relevant responses. The utilization of ChatGPT facilitates nuanced and coherent interactions, significantly improving user engagement and satisfaction.

Moreover, the AI-based chatbot holds great promise across various domains, including customer service, virtual assistance, and educational applications. Its ability to comprehend complex queries and provide accurate responses contributes to streamlining processes and enhancing productivity.

However, it's essential to acknowledge the ongoing challenges associated with AI-based chatbots, such as ensuring ethical use, addressing biases, and maintaining user privacy. Continuous refinement and adaptation of the chatbot's algorithms and training data are necessary to mitigate these challenges and ensure optimal performance.

Overall, the implementation of an AI-based chatbot leveraging ChatGPT represents a significant advancement in conversational AI technology. With further research and development, such systems have the potential to revolutionize human-computer interactions and drive innovation across various industries.

6. REFERENCES

1. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

3. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).

5. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Brew, J. (2019). HuggingFace's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771.

6. Zhang, S., Wu, Y., Du, J., Yan, Z., Huang, Y., & Wang, L. (2020). DialogGPT: Large-scale generative pre-training for conversational response generation. arXiv preprint arXiv:2004.05892.