International Journal of Research in Engineering, Science and Management
Volume 6, Issue 5, May 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Artificial Intelligence based Assistant with Specialization in a Given Language

A. Azhar Ayyash[1*], P. Dhivya Bharathi[2]

[1]*Student, Department of Information Technology, Aalim Muhammed Salegh College of Engineering, Chennai, India*

[2]*Assistant Professor, Department of Information Technology, Aalim Muhammed Salegh College of Engineering, Chennai, India*

***Abstract*: A chatbot is an artificial intelligence-based assistant with two-way interactive communication with humans. It can help them find answers to their questions, accomplish small tasks such as providing data on the users' questions, and suggest requested data with human-like Responses with the use of three AI working in synchronous order is able to create a very human-like computer assistant. These Ai could be trained in various languages and could be localized based on the data provided. It consists of an AI for Speech recognition with the help of a microphone and creating human-like responses with the use of ChatGPT API and producing them in the language of the selected requirement like English with a human-like tone and feel.**

***Keywords*: Chatbots, Open source, AI, Voice assistant, Speech recognition, Speech synthesis.**

1. Introduction

The development of artificial intelligence (AI) has led to a significant increase in productivity for individuals. Voice assistants have become increasingly popular and have made a significant impact on the way we interact with technology. One of the key features of a voice assistant is its ability to respond to voice commands, making it accessible for disabled people who may have difficulty using a keyboard or mouse. With the advancements in voice synthesis and speech recognition technology, there is now the potential for the creation of a voice assistant that is indistinguishable from a real human being. This has become possible through the use of Generative Pre-trained Transformer models (GPTs), which are powerful machine-learning algorithms that can generate human-like responses.

However, to make these assistants truly useful, they need to be able to understand and respond in the user's native language. This is where localization plays a crucial role. By localizing a voice assistant, it can be tailored to specific languages, cultures, and regions, ensuring that it can accurately interpret user commands and provide relevant responses. The use of GPTs has also opened up new possibilities for human-like responses in localized languages, making voice assistants more natural and intuitive to use. As the demand for localized voice assistants continues to grow, the development of AI will likely focus even more on language and cultural nuances, leading to even more advanced and user-friendly voice assistants.

A. *Existing System*

- Siri: A voice assistant developed by Apple, which uses natural language processing to perform various tasks on iPhone, iPad, and Mac devices.
- Google Assistant: A voice assistant developed by Google, which is available on Android and iOS devices, as well as smart speakers and smart displays. It can perform various tasks like setting reminders, sending messages, and controlling smart home devices.
- Alexa: A voice assistant developed by Amazon, which is available on Amazon Echo devices and can perform various tasks like playing music, setting reminders, and ordering products from Amazon.
- Cortana: A voice assistant developed by Microsoft, which is available on Windows devices and can perform tasks like scheduling meetings, sending emails, and providing weather updates.
- Mitsuku: A chatbot developed by Pandora bots, which uses natural language processing and machine learning to hold conversations with users on a wide range of topics.
- Xiaoice: A chatbot developed by Microsoft, which is popular in China and can hold conversations with users in Chinese on various topics, including news, entertainment, and personal advice.
- Replika: An AI chatbot that uses machine learning algorithms to create a personalized chat experience for users and help them improve their mental health and well-being.

These are a few of the real-time examples of Voice assistants and chatbot which has some kind integration of machine learning or neural network. These can mimic human interaction with a computer or smart device while providing some degree of human-like interaction. This cannot replace a human being or response like a human being

1) *Drawbacks*

- Limited Contextual Understanding: It relies on predefined responses and lacks the ability to understand complex contexts or follow-up questions as its trained data could go outdated with time.

*Corresponding author: azharayyash999@gmail.com

- Inability to handle complex requests: It sometimes struggles to handle complex requests that require multiple steps or involve multiple systems.
- Difficulty with certain accents or languages: It may have difficulty understanding certain accents or languages, limiting its effectiveness in diverse regions. This could be slang or regional language within a language.
- Lack of Emotional Intelligence: It lacks the ability to understand emotions and empathy, which can result in inappropriate responses to certain situations. (It does have some degree of emotional intelligence).
- Privacy Concerns: It collects personal data and sensitive information, there are concerns about privacy and security breaches and the use of data which is provided by the user is used to train the model.

B. Proposed Methodology

The suggested AI-based solution focuses on utilizing GPT models and localization in various languages to promote the use of AI among different generations and languages. Localization increases accessibility and usability for non-native speakers, as well as produces a diversified dataset to improve the precision and effectiveness of the GPT model. Additionally, the multilingual functionality of the system may lead to expansion in AI usage and improved training data, ultimately enhancing its accuracy and performance. This may greatly impact the development of AI-powered technology, particularly if native languages are widely used.

With various AI for language synthesis and voice recognition, it is possible to create a human-like understanding and voice creation which could create a fake person who can speak and answer like a human with the most noticeability.

2. Methodology

A. Speech Recognition

This system uses Whisper AI's open-source speech recognition system for converting speech input to text. The speech inputs Users can obtain speech from the microphone. Then The speech recognition module utilizes machine learning algorithms to convert spoken language into text by analyzing and understanding speech patterns. It includes features for noise reduction and speech enhancement and can be customized and adapted to specific speech patterns and accents. The output is text, suitable for various applications such as virtual assistants, speech-to-text dictation, and audio transcription.

B. Voice Synthesis

A Text-to-Speech (TTS) module is a software component that uses natural language processing (NLP) techniques to synthesize natural-sounding speech from written text. The output can be in various formats, such as audio files or streaming audio signals. TTS technology has advanced considerably and can produce highly realistic and expressive speech that is almost indistinguishable from human speech. TTS modules are widely used in applications like virtual assistants, in-car navigation systems, and assistive technologies for individuals with disabilities.

C. ChatGPT API

The ChatGPT API is a web-based service that allows developers to integrate natural language processing capabilities into their applications using GPT-3.5 architecture. The API is

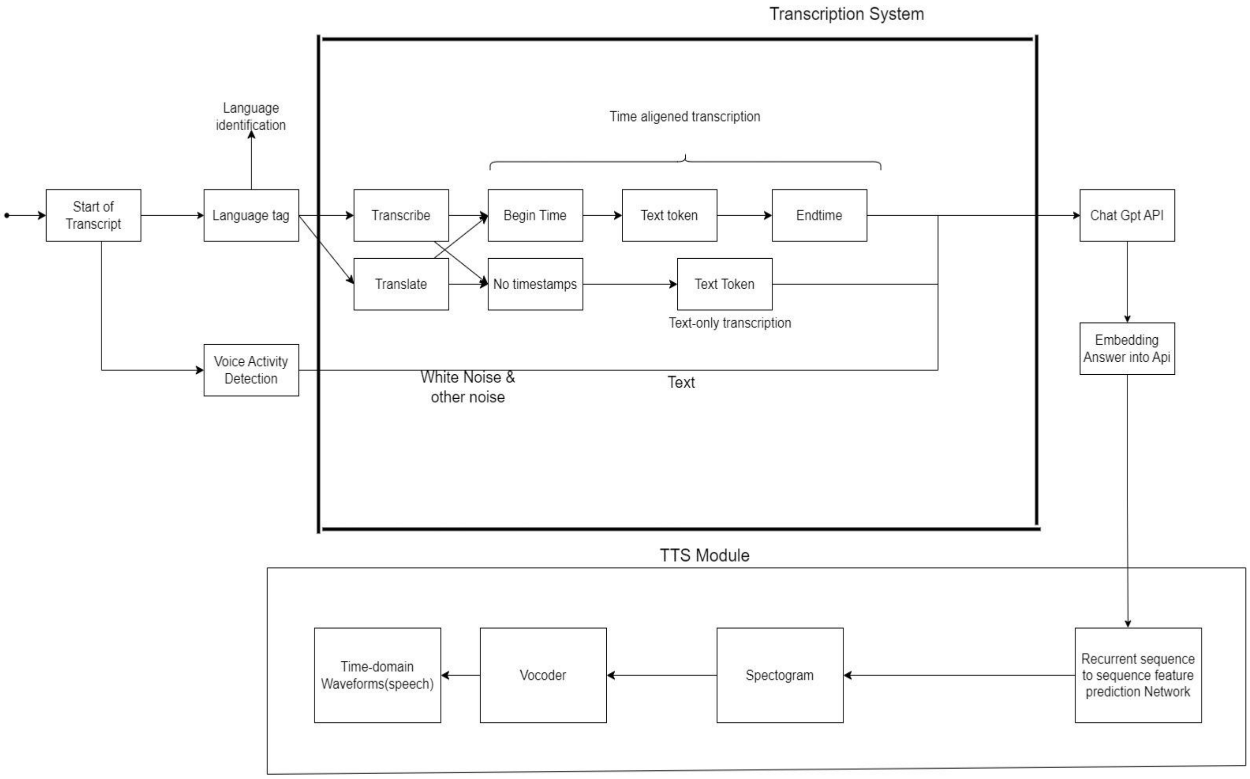

Fig. 1. Overall working of the program

Table 1
Supported languages on TTS

Local Accent	Language Code (lang)	Top-Level Domain (TLD)
English (Australia)	en	com.au
English (United Kingdom)	en	co.uk
English (United States)	en	us
English (Canada)	en	ca
English (India)	en	co.in
English (Ireland)	en	ie
English (South Africa)	en	co.za
French (Canada)	fr	ca
French (France)	fr	fr
Mandarin (China Mainland)	zh-CN	any
Mandarin (Taiwan)	zh-TW	any
Portuguese (Brazil)	pt	com.br
Portuguese (Portugal)	pt	pt
Spanish (Mexico)	es	com.mx
Spanish (Spain)	es	es
Spanish (United States)	es	us

capable of generating human-like responses in multiple languages and can understand the context of a conversation to provide relevant responses. It also allows for personalization and is scalable to handle high-traffic applications. Developers can sign up for an API key and integrate it into their applications using the provided documentation and libraries.

D. Training Sets

With the specialization in a given language, it's needed to train in various local data of the given language and slang of that language they need to be able to understand and respond in the user's native language. This is where localization plays a crucial role. By localizing a voice assistant, it can be tailored to specific languages, cultures, and regions, ensuring that it can accurately interpret user commands and provide relevant responses. The use of GPTs has also opened up new possibilities for human-like responses in localized languages, making voice assistants more natural and intuitive to use. As the demand for localized voice assistants continues to grow, the development of AI will likely focus even more on language and cultural nuances, leading to even more advanced and user-friendly voice assistants.

E. Self-supervising

The retraining of user data is an essential aspect of improving the accuracy and personalization of language in AI-based systems. As the system learns from user data, it can generate more natural and intuitive responses that better reflect the user's needs and preferences, thereby enhancing the overall user experience.

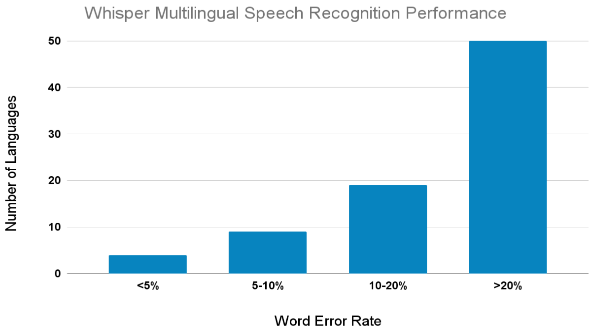

Fig. 2. Word error rate when tested on Fleurs

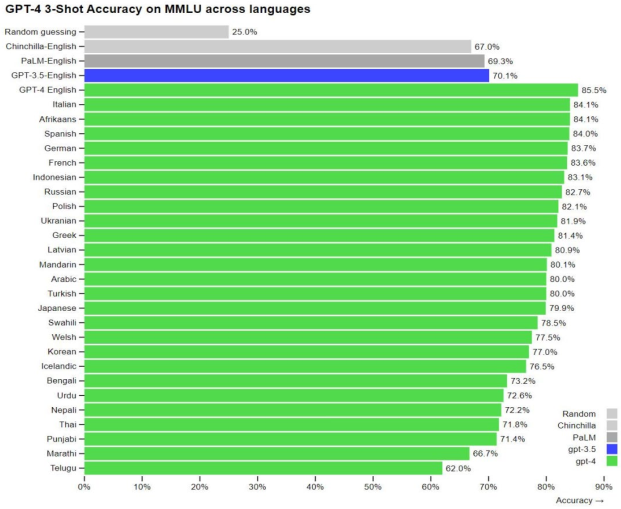

Fig. 3. GPT 3,4 accuracy on MMLU

3. Future Work

1) As we introduce the ChatGPT API, we can expect a significant improvement in various aspects of the program. With the integration of real-time voice recognition and voice identification, we can create multiple profiles to cater to the unique needs of each user. This feature will enable the program to recognize and respond to users' voices, creating a personalized and seamless experience.
2) Moreover, the ChatGPT API can integrate various smart house applications, enabling users to control all the functions of their smart homes from a single hub. This integration will simplify the process of controlling various devices, enhancing the users' convenience.
3) With the advancement of deep fake voice technology, it is possible to produce audio that mimics any person's voice so the voice produced is could on any of the user liking.
4) With the increasing user base of the application, the efficiency of the program will significantly increase. This improvement is due to the increase in data, which can help the program learn and adapt to different scenarios.
5) The TTS should support various languages in the future, including slang and different pronunciations within the same language, to cater to a diverse user base. This support

will enable the program to communicate effectively with users from different regions and backgrounds, increasing its overall functionality and usability.

6) With the advancement in voice synthesis and voice identification, it would be possible in the future to access all the languages of the civilization

4. Conclusion

In this paper, we have leveraged the power of GPT models and voice synthesis technology to create a groundbreaking approach to communication with chatbots and voice assistants. Our research has led to the development of a personal voice assistant (PVA) capable of delivering human-like speech and tone in the user's preferred language. By harnessing the capabilities of GPT models, we have been able to achieve a more natural and engaging conversational experience.

The integration of GPT models in our PVA has revolutionized the way users interact with virtual assistants. Through advanced natural language processing techniques, our PVA can generate responses that closely resemble human speech patterns, creating a seamless and immersive conversation. With the ability to adapt its responses based on user input and preferences, the PVA can personalize the interaction and provide tailored assistance to individual users.

The voice synthesis technology employed in our PVA plays a crucial role in delivering a human-like speech experience. By leveraging cutting-edge voice synthesis algorithms, we have achieved high-quality and expressive speech output that is indistinguishable from that of a real human. This breakthrough has significantly enhanced the user experience, making the interaction with the PVA more engaging, relatable, and enjoyable.

Moreover, integrating GPT models and voice synthesis technology in our PVA enables us to improve its performance and language accuracy continuously. By training the models on vast datasets and incorporating user feedback, we can refine the language generation process, expand language support, and adapt to diverse linguistic nuances. This iterative learning approach ensures that our PVA stays up-to-date and delivers the most relevant and accurate responses to user queries.

References

[1] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever Robust Speech Recognition via Large-Scale Weak Supervision

[2] H. Larochelle and M. Ranzato and R. Hadsell and M.F. Balcan and H. Lin Advances in Neural Information Processing Systems 33 (NeurIPS 2020).

[3] Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, Paul F. Christiano Learning to summarize with human feedback.

[4] Leo Gao, John Schulman, Jacob Hilton Scaling Laws for Reward Model Overoptimization.

[5] Miles Brundage et al., Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims.

[6] Josh A. Goldstein, Girish Sastry, Micah Musser, Renee DiResta, Matthew Gentzel, Katerina Sedova, Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations.

[7] Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, Michael Auli wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations.

[8] Daniel Galvez, Greg Diamos, Juan Ciro, Juan Felipe Cerón, Keith Achorn, Anjali Gopi, David Kanter, Maximilian Lam, Mark Mazumder, Vijay Janapa Reddi, The People's Speech: A Large-Scale Diverse English Speech Recognition Dataset for Commercial Usage.

[9] Yu Zhang et al., BigSSL: Exploring the Frontier of Large-Scale Semi-Supervised Learning for Automatic Speech Recognition.

[10] William Chan, Daniel Park, Chris Lee, Yu Zhang, Quoc Le, Mohammad Norouzi SpeechStew: Simply Mix All Available Speech Recognition Data to Train One Large Neural Network.

[11] Cortana Intelligence, Google Assistant, Apple Siri.

[12] Hill, J., Ford, W.R. and Farreras, I.G., 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior, 49, pp. 245-250.

[13] K. Noda, H. Arie, Y. Suga, T. Ogata, Multimodal integration learning of robot behavior using deep neural networks, Elsevier: Robotics and Autonomous Systems, 2014.

[14] "CMUSphnix Basic concepts of speech - Speech Recognition process". http://cmusphinx.sourceforge.netlwiki/tutorialconcepts

[15] Huang, J., Zhou, M. and Yang, D., 2007, January. Extracting Chatbot Knowledge from Online Discussion Forums. In IJCAI, vol. 7, pp. 423-428.

[16] Thakur, N., Hiwrale, A., Selote, S., Shinde, A. and Mahakalkar, N., Artificially Intelligent Chatbot.

[17] Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in Human-Computer Interaction. In 5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) (pp. 79-84).

[18] Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning & Technology.

[19] Mozilla's large repository of voice data will shape the future of machine learning. https://opensource.com/article/18/4/common-voice

[20] Hill, I. (1983). "Natural language versus computer language." In M. Sime and M. Coombs (Eds.) Designing for Human-Computer Communication. Academic Press.