iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49536

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

Voice Assistant based on GPT-3 and Whisper

V. Sai Vivek[1], P. Prudhvi Raj[2], K. Santhosh[3], Dr. Vithya Ganesan[4]
[1, 2, 3, 4]*K L University*

Abstract: ***A voice assistant based on GPT-3 (Generative Pre-Trained Transformer 3) and Whisper, a secure multi-party computation technology, can offer an enhanced level of privacy and security to users. By leveraging Whisper, GPT-3 can securely process user data without compromising privacy, ensuring that sensitive information remains protected. This combination also enables the development of decentralized voice assistants, which can operate without relying on a central server or data centre, further enhancing privacy and security. With this technology, users can interact with the voice assistant confidently, knowing that their data is secure and private. Additionally, the natural language processing capabilities of GPT-3 make the interaction with the voice assistant more intuitive and seamless, allowing users to complete tasks quickly and efficiently. Overall, a voice assistant based on GPT-3 and Whisper offers a powerful and secure tool for individuals and organizations that prioritize privacy and security.***
Keywords: Voice Assistant, GPT-3, Large Language Model, Whisper.

I. INTRODUCTION

Voice assistants have become increasingly popular in recent years, as they provide a convenient and intuitive way for users to interact with their devices. These computer programs are designed to recognize and respond to voice commands, allowing users to perform a wide range of tasks, from setting reminders and searching the web, to controlling smart home devices and playing music. However, the focus of this research paper will be on a specific type of voice assistant that is designed specifically for answering questions. The voice assistant architecture described in this research paper utilizes Open AI's open-source speech-to-text application, the GPT-3 language model, and the pyttsx3 text-to-speech conversion library to understand, process, and provide answers to user queries. This architecture provides a simple and effective solution for understanding and responding to user questions, making it a valuable tool for individuals looking for quick and accurate information.

II. WORKING

This project aims to develop an AI voice assistant that can understand and respond to voice commands in a natural and human-like manner. The voice assistant utilizes three main technologies: OpenAI's Whisper Speech-to-Text model, GPT-3's textdavinci-003 API, and pyttsx3's Text-to-Speech technology. The Whisper model is an open-source Speech-to-Text model that functions similarly to a human ear, transcribing voice input into text. The transcribed text is then sent to GPT-3's text-davinci-003 API, which generates a response based on the prompt received. This response is then converted to human-like speech using pyttsx3's Text-to-Speech technology. In this way, Whisper acts as the "ear" of the voice assistant, GPT-3 acts as the "brain," and pyttsx3 acts as the "mouth" producing speech. This integration allows for more natural and human-like interaction with the voice assistant. The end result is an AI voice assistant that can understand and respond to voice commands in a natural and human-like manner.

A. Audio Data Collection

The first step in building a voice assistant that is capable of answering questions is to collect audio data from the user's voice. This is done using Open AI's open-source speech-to-text application, which acts as the "ear" of the voice assistant. The speech-to-text application uses state-of-the-art machine learning algorithms to convert the user's voice into text, providing an accurate representation of the user's query.

B. Processing with GPT-3

Once the audio data is collected, it is then sent to GPT-3 for processing. GPT-3 is a state-of-the-art language model developed by OpenAI that has been trained on a massive corpus of text data. It uses advanced artificial intelligence algorithms to understand the user's query and generate a response. GPT-3 acts as the "brain" of the voice assistant, providing the intelligence and reasoning necessary to understand and respond to user queries.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

C. Outputting the Response

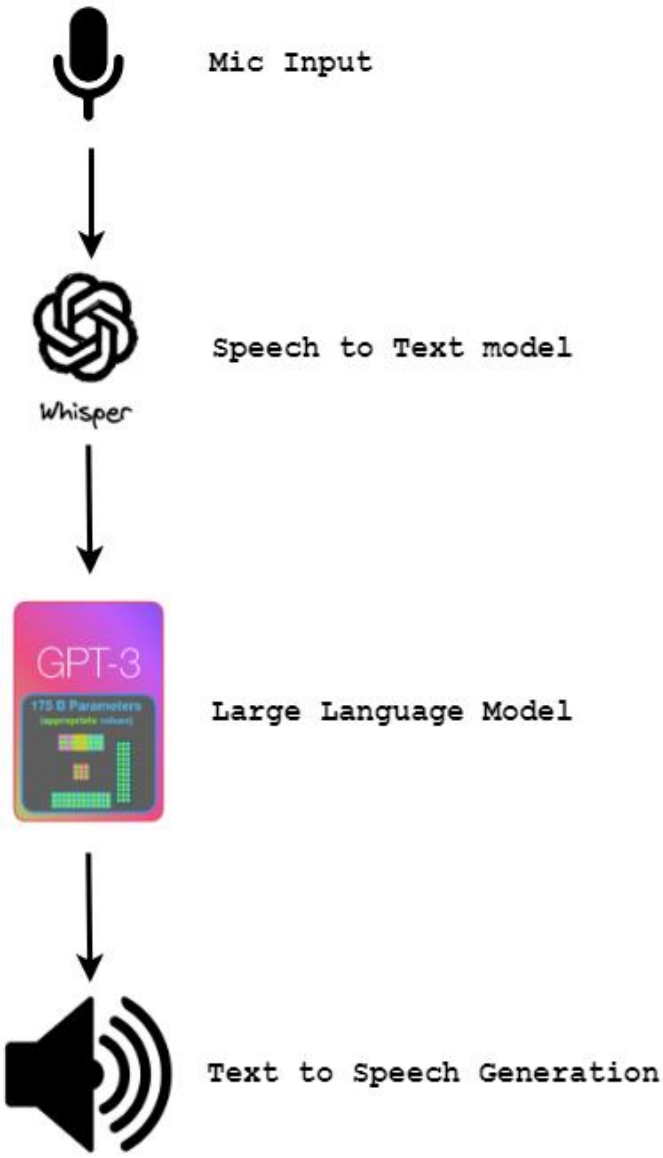
Once the response has been generated by GPT-3, it is then converted to speech using the pyttsx3 text-to-speech conversion library. Pyttsx3 is a python library that allows for the conversion of text to speech, and can be easily integrated into the voice assistant architecture. The response generated by GPT-3 is read out loud to the user, providing them with a convenient and intuitive way to receive information.

III. VOICE ASSISTANT ARCHITECTURE

The voice assistant architecture described in this research paper offers several advantages over other voice assistant technologies. The use of Open AI's open-source speech-to-text application and the GPT-3 language model provides the voice assistant with state-of-the-art machine learning algorithms and a vast corpus of text data, allowing it to accurately understand and respond to a wide range of user queries.

Additionally, the use of the pyttsx3 text-to-speech conversion library provides the voice assistant with a flexible and easily-integrated solution for outputting the response. This allows for quick and easy development and deployment of the voice assistant, providing users with a convenient and intuitive way to receive information.

IV. BLOCK DIAGRAM



©IJRASET: All Rights are Reserved

V. APPLICATIONS AND ADVANTAGES

One of the key benefits of this voice assistant architecture is its ability to scale and adapt to changing user needs. As the GPT-3 language model continues to be trained on larger and more diverse corpora of text data, the voice assistant will become increasingly capable of understanding and responding to a wider range of questions. This will allow the voice assistant to remain relevant and useful to users over time, even as their needs and interests change.

It is important to note that this voice assistant architecture can also be applied to a wide range of use cases and industries. For example, in the education sector, the voice assistant can be used to provide students with quick and easy access to information, enabling them to quickly find answers to their questions and further their understanding of a particular subject. In healthcare, the voice assistant can be used to provide patients with information about their symptoms, conditions, and treatments, helping them to make informed decisions about their health.

In the business world, the voice assistant can be used to automate customer service and support, providing customers with quick and accurate answers to their questions and reducing the workload on human customer service representatives. This can result in increased customer satisfaction and reduced costs for the business.

Automating customer service is a critical aspect of many businesses, as it enables organizations to provide quick and efficient support to their customers, while reducing the workload on human customer service representatives. The voice assistant architecture described in this research paper can play a key role in automating customer service by providing customers with quick and accurate answers to their questions.

By integrating the voice assistant into a customer service platform, businesses can provide their customers with a self-service option for obtaining information about products, services, and policies.

This can help to reduce the volume of inbound customer support calls and emails, freeing up customer service representatives to focus on more complex issues.

The voice assistant can understand and respond to customer questions in natural language, making it easy for customers to use. It can also be configured to provide information on a wide range of topics, allowing it to handle a wide variety of customer queries. This can help to reduce the need for customers to wait for a response from a customer service representative and can also improve customer satisfaction by providing them with the information they need quickly and easily.

In addition to providing quick and accurate answers to customer questions, the voice assistant can also be used to gather customer feedback and insights. This can help businesses to better understand the needs and concerns of their customers, and to make informed decisions about product and service offerings.

Another key benefit of the voice assistant for customer service is its ability to handle large volumes of interactions. As the voice assistant is powered by the GPT-3 language model, it is able to handle a large number of interactions simultaneously, making it an ideal solution for businesses that receive a high volume of customer support queries.

Finally, the voice assistant can be easily integrated into existing customer service platforms, making it a cost-effective and scalable solution for automating customer service. This can help businesses to reduce the costs associated with customer support, while still providing their customers with the high-quality support they need.

Another important advantage of this voice assistant architecture is its ability to provide accurate and reliable information to users. By utilizing the GPT-3 language model, the voice assistant is able to understand and respond to user queries based on a vast corpus of text data, providing users with trustworthy and relevant information. This is particularly important in contexts where accuracy and reliability are critical, such as in healthcare or financial applications.

VI. FUTURE DEVELOPMENT

In terms of the future of voice assistant technology, there are several exciting developments to look forward to. For example, as the GPT-3 language model continues to be trained on larger and more diverse corpora of text data, the accuracy and relevance of the voice assistant's responses is likely to improve. Additionally, as natural language processing technology continues to advance, the voice assistant is likely to become increasingly capable of understanding complex user queries and providing accurate and relevant responses.

Another potential future development is the integration of machine learning algorithms into the voice assistant architecture. This could enable the voice assistant to learn from its interactions with users, becoming more accurate and relevant over time. This could also open up new possibilities for personalized and context-aware interactions, allowing the voice assistant to provide users with tailored and relevant information.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VII. PRIVACY & SECURITY

In terms of privacy and security, this voice assistant architecture also offers several advantages. By utilizing the API call to send the user's audio data to the GPT-3 language model, the voice assistant is able to keep user data secure and protected. Additionally, the use of an open-source speech-to-text application and text-to-speech conversion library allows for greater transparency and control over the data being used by the voice assistant. This helps to ensure that user data is being used in a responsible and ethical manner.

VIII. LIMITATIONS

In terms of potential limitations, there are several factors to consider when using this voice assistant architecture. Firstly, the GPT-3 language model is only as good as the data it has been trained on, and may not always provide accurate or relevant responses to user queries. Additionally, the speech-to-text application may struggle to accurately transcribe audio data in noisy or challenging environments. Finally, the text-to-speech conversion library may not provide a high-quality output in all cases, particularly in terms of pronunciation or intonation.

To overcome these limitations, it may be necessary to fine-tune the voice assistant architecture by incorporating additional machine learning algorithms, or by training the GPT-3 language model on more diverse corpora of text data. Additionally, it may be necessary to develop and integrate additional speech-to-text and text-to-speech applications to improve the accuracy and reliability of the voice assistant.

IX. RESULTS

```
AI Voice Assistant is listening

User input receiving completed

Q:  How many layers are there in a neural network?
A: There is no single answer to this question as the number of layers in a neural
 network can vary depending on the type of neural network and the specific applic
ation. Generally speaking, neural networks can have anywhere from one to hundreds
 of layers.
```

```
AI Voice Assistant is listening

User input receiving completed

Q:  What is the capital of India?
A: The capital of India is New Delhi.
```

```
AI Voice Assistant is listening

User input receiving completed

Q:  What is the distance between the sun and the earth?
A: The average distance between the Sun and the Earth is about 93 million miles (149.6 million ki
lometers).
```

```
AI Voice Assistant is listening

User input receiving completed

Q:  And what principle does rocket's work?
A: A rocket works on the principle of Newton's Third Law of Motion, which states that for every a
ction, there is an equal and opposite reaction. This means that when a rocket expels a large amou
nt of exhaust gas in one direction, the rocket will be propelled in the opposite direction.
```

Subjective Q&A:

```
AI Voice Assistant is listening

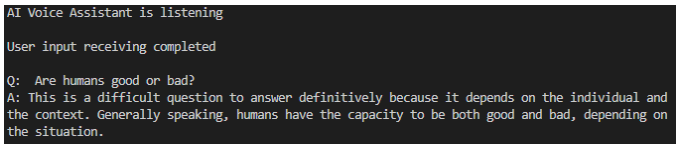
User input receiving completed

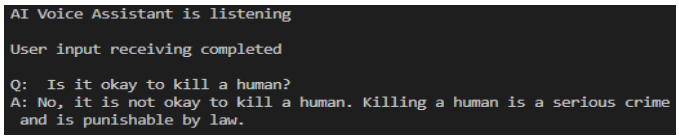
Q:  Who is a good president? Joe Biden or Donald Trump?
A: The answer to this question is subjective and depends on the individual's opinion.
```

```
AI Voice Assistant is listening

User input receiving completed

Q:  Will USA go to war on Russia?
A: No. The United States and Russia are not currently at war, and there is no indication that
 either country is planning to go to war with the other.
```



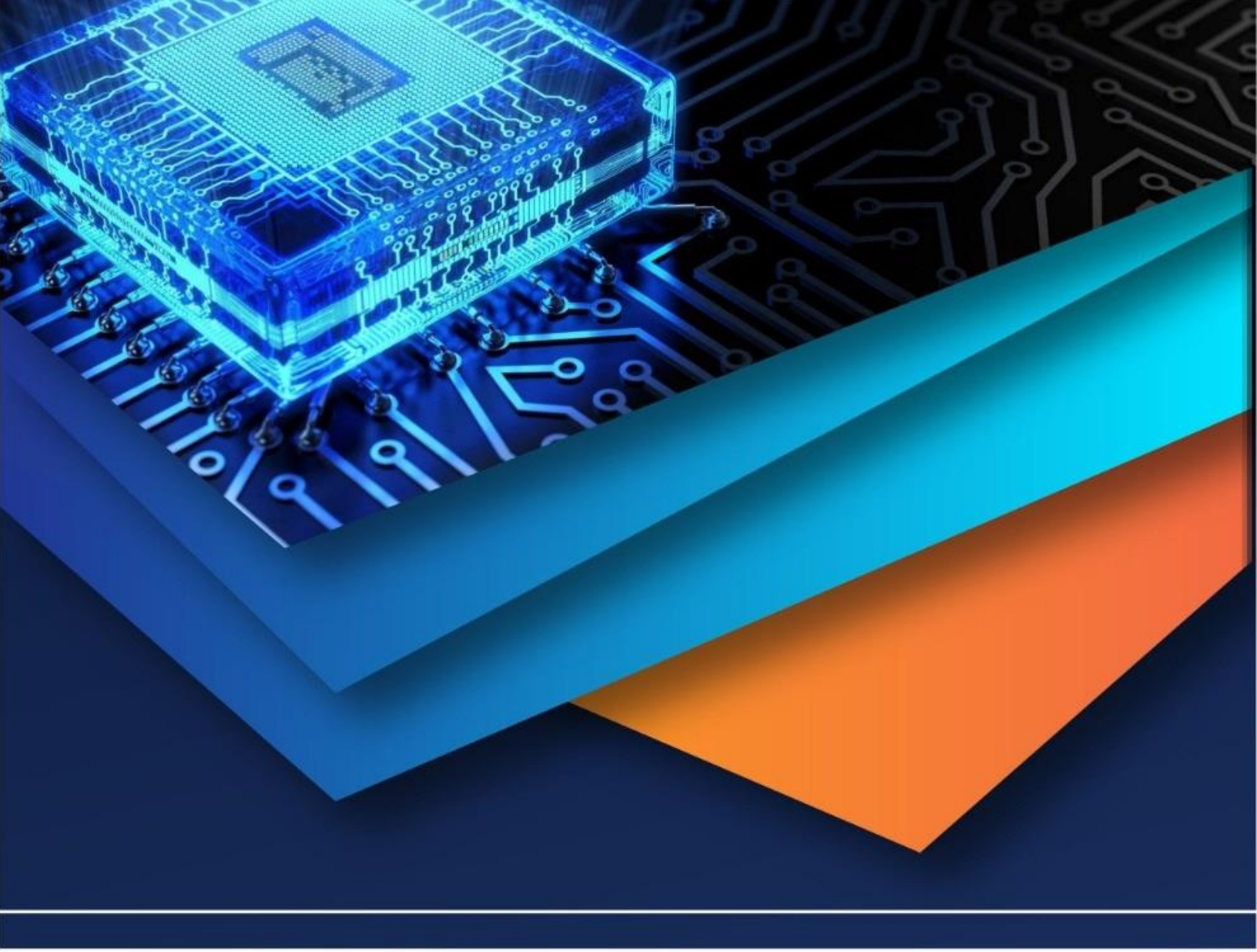


X. CONCLUSION

In conclusion, the voice assistant architecture described in this research paper provides a simple and effective solution for understanding and responding to user questions. By utilizing Open AI's open-source speech-to-text application, the GPT-3 language model, and the pyttsx3 text-to-speech conversion library, this architecture provides a flexible and easily-integrated solution for developing a voice assistant that can accurately understand and respond to user queries. While there are some limitations to consider, the potential benefits of this voice assistant architecture make it a valuable tool for individuals and organizations looking for a convenient and accurate way to receive information.

REFERENCES

[1] Voice Assistants and Smart Speakers in Everyday Life and in Education, George TERZOPOULOS, Maya SATRATZEMI, 2020
[2] Language Models are Unsupervised Multitask Learners, Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever
[3] Language Models are Few-Shot Learners, Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, 2020
[4] Robust Speech Recognition via Large-Scale Weak Supervision, b Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskevery, 2022



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)