Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, May : 2024

VOICE DRIVEN CHATGPT

[1]Chapala Jhansi Durga Sree,[1]Karri Visalakshi,[1]Ithireddy Indu kumari,[1]Adapa Ramya, [1]Monisha Barua, [2]Mrs. K. Therissa
[2]Assistant Professor, [1,2] Department of Electrical and Electronics Engineering Vignan's institute of Engineering for women, Visakhapatnam, Andhra Pradesh,India

ABSTRACT:
Now a days chatGpt is the most trending AI tool that utilizes natural language processing (NLP) to understand and respond to user queries in real time, ensuring a smooth conversational flow, all within a text-based interface. To augment user experience with voice interactions developed a voice-driven ChatGpt interface. Voice interactions are becoming important in user experience design (UED), as they enable users to communicate using natural language and speech, eliminating the need for hand free communication. The voice is taken as input for this voice-driven ChatGpt interface and it is converted into text using API keys. Subsequently, this textual input undergoes processing via the ChatGPT interface, ultimately resulting in the delivery of responses in the form of voice output, catering in need of speaking and listening instead of text. The goal is to enhance user experience while maintaining efficiency through this interface.
Keywords: NLP, UED, Text -Based Interfaces, ChatGPT Interface, API Keys
Abbreviations: Natural Language Processing (NLP), User Experience Design (UED), Application Programming Interface (API), Artificial Intelligent (AI).

INTRODUCTION:

User's concerns play a crucial role in the invention of technology because they provide valuable insights into what features, functionalities, and improvements are needed to address real-world problems and enhance user experiences. By listening to and addressing these concerns, guides innovators to create solutions that are more relevant, effective, and user-friendly, ultimately leading to greater adoption and success of the technology in the market. Artificial Intelligence (AI) represents the pinnacle of technological innovation because they learn unobtrusively from our interaction behavior, store every interaction and can react adaptively and even make predictions about our next behaviour. They acquire some degree of human-like cognitive, self-executing, and self-adaptive capabilities and autonomy, and produce unexpected outputs that require non-deterministic interactions. One important technological advancement that has the power to drastically alter whole markets, industries, commercial operations, and business models is artificial intelligence (AI). These artificial neural network-based computer models for language processing, which are assisted by AI, are frequently used to analyze vast volumes of data and complicated language. This is because artificial neural networks facilitate the effective translation of language into mathematical parameter. This topic has gained significant attention since November 2022, when ChatGPT, an AI-supported computer model for language processing, was released by the US business Open AI. In a matter of days, ChatGPT reached millions of users worldwide. The launch of ChatGPT-4, which can handle text and picture input, including written documents, images, diagrams, or screenshots, came next. The manufacturer claims that ChatGPT-4 outperforms humans in a variety of professional and academic benchmarks. Interms of education, these technologies could help students think critically about scientific methods, improve their messaging, have messages edited, or even serve as tutors for test preparation. But these advantages come with drawbacks

as well, such as false information, security issues, and a lack of scientific rigor. Voice assistants are AI-powered digital tools that understand and respond to spoken commands. Voice-mediated service experience is made possible by these voice-based interfaces, which integrate natural language processing and automated speech recognition. They have gained popularity for their convenience, enabling users to perform tasks like setting reminders, playing music, and controlling smart devices hands-free. This indicates that consumers are becoming more accustomed to interacting through voice-based interfaces, and that businesses are expanding the use of voice-mediated service channels across touch points. The present voice assistants are designed to perform a wide range of tasks such as setting reminders, sending messages, playing music, controlling smart home devices, and offering recommendations based on user commands and to provide personalized recommendations. Even though we have so many text-based interfaces, ChatGPT's interface and its features are indeed designed to provide a conversational and engaging experience for users which makes its raising popularity among students. Its ability to understand context, generate coherent responses, and offer a wide range of information makes it a valuable tool for various purposes, from learning and exploring topics to getting personalized assistance and recommendations. The Voice driven ChatGPT interface includes the features that ChatGPT interface is providing which enhances the capabilities of the voice assistance according to the user and helps to make studying and learning more accessible, interactive, and efficient by leveraging the power of voice technology and AI-driven capabilities.

LITERATURESURVEY:

Technologies are advancing rapidly in a blink of eye irrespective of the field. New innovations are constantly being developed but not every0 new technology can lives up to people's expectations and one of the main concern for the failure is mostly because they fail to keep up with changing user needs. Even the successful technologies require continual modification and adaptation to remain relevant in rapidly evolving markets and societal trends. Conversational AI models are epoch-making trend in the field of natural language processing (NLP) and artificial intelligence (AI) which has gained popularity for their ability to generate human like text and engage in realistic conversations. Voice based interfaces have renown despite the rise of the text based interfaces due to their advanced abilities and capabilities.The voice driven user interfaces are typically developed using various AI libraries such as Tensor Flow, Py Torch, or specialized speech recognition libraries. These interfaces utilize machine learning models for tasks but their capabilities are often limited to specific cases or functionalities defined by their developers which requires explicit programming for each interaction scenario and may not have the ability to engage users in open-ended conversations. "Personas for Artificial Intelligence (AI) an Open-Source Toolbox" by Andreas Holzinger, Michaela Kargl, Bettina Kipperer, Peter Regitnig, Markus Plass and Heimo Muller (2022): It summarizes that AI systems exhibit various capabilities ranging from human-like behavior in communication to providing autonomy and contextual understanding. They can solve complex classification problems, enable intelligent interaction, and generate non-deterministic output like co-creating music. Additionally, AI systems can collaborate with humans as teammates, augment human intelligence, and often operate as opaque "black box" systems, making their inner workings challenging to understand or verify. "Research and Analysis on Voice Based System with Machine Learning "by Gopal Vishwas Patil and Vidya Dhamdhere (2022): For automation of the repetitive tasks, chat bots have transitioned from text-based to voice-based interfaces incorporating hot word detection like Alexa for activation which starts Tensor flow JS model for further speech recognition. It enhances user involvement through seamless interaction, resulting in optimized task execution efficiency. To achieve this python based API keys for converting the audio into text through

speech recognition algorithms. NLP based interpreter recognize text with the actual data which was predefined and then the parser gathers and structures the information to generate a response suitable for the user interface, ensuring consistent interaction with the user. The python API keys requires more development effort and used if there is a need to integrate with a variety of third party services. The selection of API keys depends on the functionalities of the project based on the parameters such a cost, security, integration with existing technology and many more factors. The security is totally depends on the developer. "A Review of Subjective Scales Measuring the User Experience of Voice Assistants" by Lawal Ibrahim Dutsinma Faruk, Mohammad Dawood Babakerkhell, Pornchai Mongkolnam, Vithida Chongsuphajaisiddhi, Suree Funilkul and Debajyoti pal (2024): No matter what the outcome of the product, the final decision maker is user so the first priority is to look into user needs which helps in the advancing the needs of user. This document highlights the changing facets of user experience in Voice Assistant (VA) research. Stating integrating human-like qualities or behaviours to VAs, which pocess distinct traits, behaviours that shape their interactions with users, aiming to enhance engagement and usability. These elements contribute to a more immersive and satisfying user experience. "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development" by Tianyu Wu , Shizhu He , Jingping Liu , Siqi Sun , Kang Liu , Qing-Long Han , Fellow, IEEE, and Yang Tang(2023):This provides an outline of ChatGPT, an advanced AI agent, discussing its predecessor, strengths, limitations, social impact, and potential future development. It explores core techniques such as large-scale language models and reinforcement learning, highlighting the emergence phenomenon in AI. "Unleashing the Potential of Conversational AI: Amplifying Chat-GPT's Capabilities and Tackling Technical Hurdles "by Vikas Hassija, Arja Chakrabarti, Anushka Singh, Vinay Chamolaand Biplab Sikdar (2023):This signifies that large language models (LLM) are overtaking natural language processing(NLP) but facing some challenges in fixing the problems. It provides the capabilities, limitations of LLM in voice companions, and underlying techniques such as deep learning and reinforcement learning. It gives a brief of basic preparation of training data and the existing technological knowledge for intercepting the decision making. Overall, the conclusion underscores the significant advancements made in Conversational AI and the vital role played by technologies in shaping the future of human-machine interaction. Though there are different API keys and AI interfaces which can be opted based on the functionality and each poses their unique capabilities and limitations. In voice driven chatGPT interface the google api keys are used because as it provides more security and easy integration when compared to python based api keys. Whereas, Chat GPT interfaces utilize Open AI's models to generate human like response allowing and engage in open- ended conversations.

EXISTINGMETHOD:

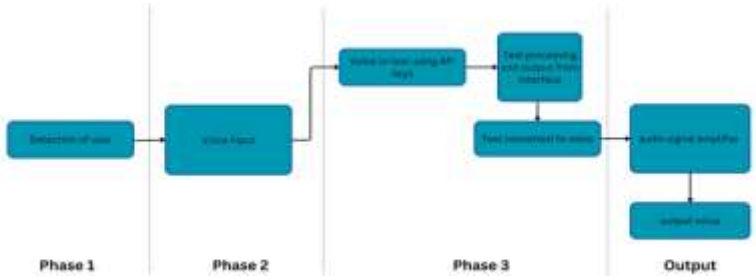
Before Voice-Driven ChatGPT came along, interacting with AI models like ChatGPT usually meant typing out your questions or commands. While this worked fine for some people, it wasn't very accessible for everyone, especially those who preferred speaking or had trouble typing. Plus, typing out everything could take a lot of time and effort, making the whole experience less engaging and efficient.
The existing system of Voice-Driven ChatGPT Interface changes all that by letting users talk to the AI model instead of typing. This makes it much easier and more natural to get things done, whether you're asking for information, getting help with tasks, or just having a friendly conversation. By introducing voice commands, we're making AI technology more inclusive and user-friendly, catering to diverse preferences and needs. The transition from typing to voice interaction with the Voice-Driven ChatGPT Interface marks a significant improvement in user experience. By allowing users to speak directly to the AI model, we eliminate barriers for those who may struggle with typing, such as individuals with

mobility impairments or those who are visually impaired. This enhancement not only promotes accessibility but also fosters a more natural and intuitive mode of communication, mimicking real-life conversations.

Moreover, the integration of voice commands enhances efficiency and engagement in various tasks. Users can effortlessly navigate through tasks, access information, and execute commands with simple voice prompts, reducing the time and effort required compared to traditional typing methods. This seamless interaction empowers users to accomplish tasks more effectively, whether they are seeking quick answers, managing their schedules, or seeking personalized recommendations. In essence, the Voice-Driven ChatGPT Interface represents a leap forward in human-computer interaction, ushering in a new era of accessibility, efficiency, and user-centric design in AI technology.

METHOLODGY:

The whole process is actually divided into two segments they are converting speech to text and converting text to speech: For Converting speech to text and text to speech, we generally use API keys. These API's cover various areas such as computing, storage, machine learning, data analytics, and more. Google Cloud API provides extensive documentation ,tutorials, and support resources to help developers get started and make the most out of the platform at present there are several API keys which purposes different category and some of the available API keys for converting speech to text includes Google Cloud Speech-to-Text API key, IBM Watson Speech to Text API key, Microsoft Azure Cognitive Services Speech API key, and Amazon Transcribe API key. Among the various API's which are available, offers different considerations such as pricing, language support, accuracy, and features, tailored to meet the diverse needs of users. Google Cloud's Text-to-Speech API offers an extensive language support, robust customization options, seamless integration, accuracy and scalability, competitive cost all these elements makes to choose the google cloud API key. Taking all the factors into account ,Opting for Google Cloud API keys is the optimal choice for our project



Structural Representation of Voice Driven ChatGPT Interface

5. ANALYSIS:

India boasts a diverse population, with over 22 official languages. Despite this linguistic diversity, ChatGPT has gained significant traction in the country .India is the second country after United States regarding ChatGPT users, with 8.37% of all users. ChatGPT is currently utilized in India primarily for text-based applications, including customer support and content generation.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, May : 2024

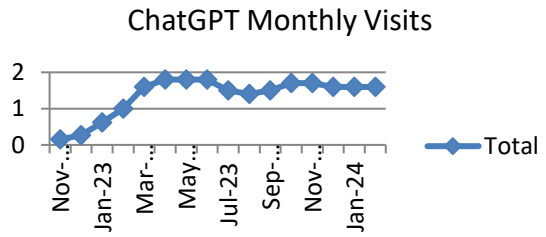


Fig1.ChatGPT monthly visits progressed over time

ChatGPT is designed with user experience in mind, aiming for simplicity and natural interaction. By mimicking the format of familiar messaging apps, users feel comfortable engaging with it. This approach helps bridge the gap between humans and technology, fostering trust and acceptance of AI-powered chat interfaces. One key feature of ChatGPT is its ability to adapt its responses to match the tone and sentiment of user input. It feels more like chatting with a real person because it's powered by a massive pre-trained language model that learned from vast amounts of data. This ensures that interactions feel personalized and relevant, enhancing the overall user experience. Moreover, ChatGPT is equipped with content filters to uphold ethical guidelines and community standards. These filters prevent the generation of responses that could be inappropriate, offensive, or harmful. By incorporating these safeguards, ChatGPT promotes responsible use of AI technology and maintains a safe environment for users. Overall, ChatGPT's focus on simplicity, natural language interaction, and ethical guidelines makes it a user-friendly and trustworthy platform for powering intelligent chatbots across various domains.

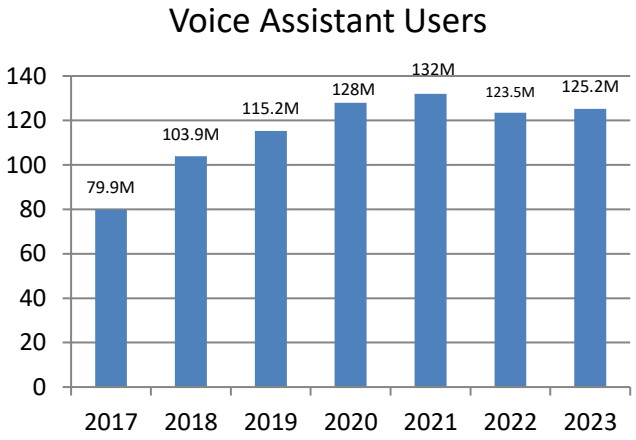


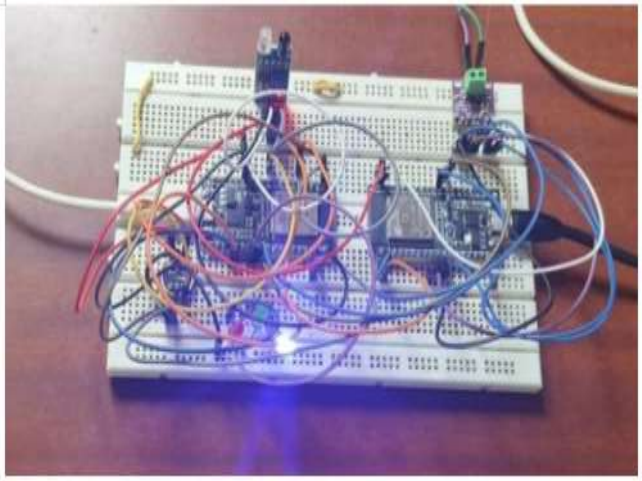
Fig2. ChatGPT monthly visits progressed over time

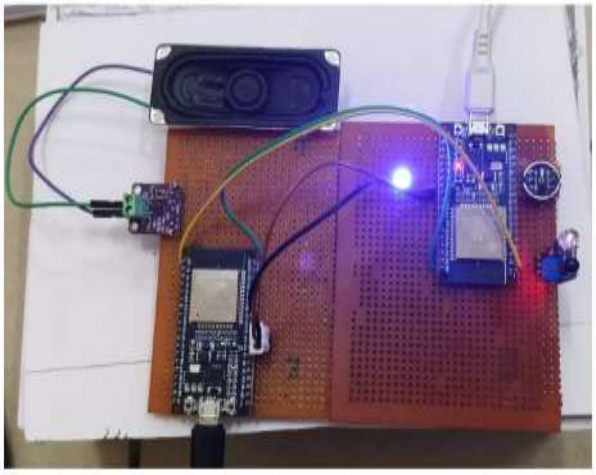
Voice based interfaces have renown despite the rise of the text based interfaces due to their advanced abilities and capabilities. It helps to make studying and learning more accessible, interactive, and efficient by leveraging the power of voice technology. AI-driven interfaces can be designed to cater to users with diverse needs and abilities, such as providing support for individuals with visual or auditory impairments. Features like voice control, text-to-speech, and customizable interfaces can make these systems more accessible to a wider range of users, ensuring that everyone can benefit from their capabilities. The main purpose of Voice-Driven ChatGPT is to enable seamless and intuitive interaction between users and the ChatGPT model through spoken language. This aims to enhance accessibility, engagement, and usability in various applications and domains. Additionally, Voice-Driven ChatGPT

seeks to cater to diverse user preferences, including those with visual impairments, by allowing natural language communication and providing features such as real-time speech recognition, natural language understanding, and contextual comprehension. It serves as a virtual assistant, enhances customer support, supports education, aids content creation, improves accessibility, and provides personalized recommendations. Furthermore, Voice-Driven ChatGPT serves as a valuable study assistance tool for students. By leveraging its capabilities in natural language understanding and contextual comprehension, students can use the platform to ask questions, receive explanations, and access educational resources in a conversational manner. In addition to its objectives, Voice-Driven ChatGPT also aims to benefit blind people by providing enhanced accessibility through spoken language interaction. Overall, it enables users to access information, receive assistance, and engage in conversations seamlessly using voice commands. those with visual impairments, by allowing natural language communication and providing features such as real-time speech recognition, natural language understanding, and contextual comprehension. It serves as a virtual assistant, enhances customer support, supports education, aids content creation, improves accessibility, and provides personalized recommendations. Furthermore, Voice-Driven ChatGPT serves as a valuable study assistance tool for students. By leveraging its capabilities in natural language understanding and contextual comprehension, students can use the platform to ask questions, receive explanations, and access educational resources in a conversational manner. In addition to its objectives, Voice-Driven ChatGPT also aims to benefit blind people by providing enhanced accessibility through spoken language interaction. Overall, it enables users to access information, receive assistance, and engage in conversations seamlessly using voice commands.

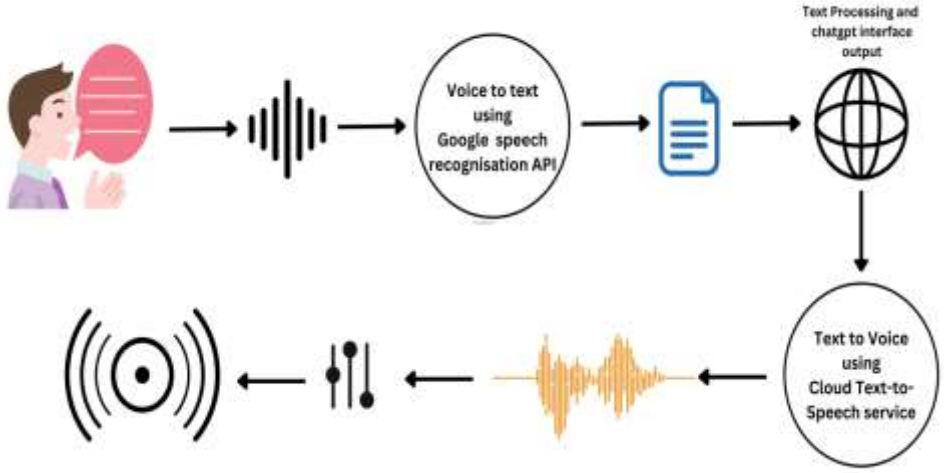
RESULTANALYSIS:

The voice input is captured by the mic and processed to the Esp 32 and the processed voice is converted into text and later is used by cloud speech client. The text is processed and the output from cloud speech client is in the form of text. The text is later processed into voice. At the last stage the voice output is amplified and from speaker the users can hear.





Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, May : 2024



CONCLUSION:

Voice assistants have become indispensable tools in our daily lives, offering unparalleled convenience and efficiency through their ability to understand and respond to spoken commands.The rise of conversational AI models like ChatGPT has further elevated the appeal of voice assistants, particularly among students. ChatGPT interface stands out for its conversational and engaging experience, making it increasingly popular among users seeking interactive and personalized assistance through text based interface. This voice-based ChatGPT interface leverages ChatGPT's remarkable ability to understand context, generate coherent responses, and offer a vast array of information and enabling it to produce voice output seamlessly. it will be a valuable companion for learning, exploration, and accessing personalized recommendations. As technology continues to evolve, the integration of conversational AI into voice assistants promises to further enhance their functionality and usability.

FUTURE SCOPE:

The future scope of the present systemholds immense potential for innovation across various industries and applications. As technology continues to advance interactions with these systems will become smoother, easier, and much more efficient. One key area of innovation lies in the enhancement of multilingual capabilities. As the world becomes increasingly interconnected, breaking down language barriers becomes essential for communication and collaboration. This would greatly improve diversity and inclusivity, allowing users from different linguistic backgrounds to interact effortlessly.

By leveraging past interactions and user preferences, AI-driven interfaces can tailor responses and recommendations according to individual preferences. Such customization enhances user experience and efficiency, as the system becomes more aligned with each user's unique needs and preferences over time. It holds promise for advancing communication, collaboration, and accessibility across various domains by incorporating mentioned features which facilitates seamless, efficient, and inclusive interactions in an increasingly interconnected world.

7. REFERENCES:

[1] Gopal Vishwas Pati and Vidya Dhamdhere ,"Research and Analysis on Voice Based System with Machine Learning", 2022 10th IEEE International Conference on Emerging Trends in Engineering &Technology-Signal and Information Processing (ICETET-SIP- 22). DOI: 10.1109/ICETET-SIP-2254415.2022.9791498.

[2] Andreas Holzinger, Michaela Kargl, Bettina Kipperer, Peter Regitnig, Markus Plass and Heimomuller "Personas for Artificial Intelligence (AI) an Open Source Toolbox" (2022). DOI: 10.1109/ACCESS.2022.3154776.

[3] Vikas Hassija, Arjab Chakrabarti, Anushka Singh, Vinay Chamola and Biplab Sikdar ,"Unleashing the Potential of Conversational AI: Amplifying Chat-GPT's Capabilities and Tackling Technical Hurdles " (2023).DOI: 10.1109/ACCESS.2023.3339553.

[4] Lawalibrahim Dutsinma Faruk, Mohammad Dawood Babakerkhell, Pornchai Mongkolnam, Vithida Chongsuphajaisiddhi, Suree Funilkul and Debajyoti Pal, "A Review of Subjective Scales Measuring the User Experience of Voice Assistants" (2024).DOI: 10.1109/ACCESS.2024.3358423.

[5] Tianyu Wu , Shizhu He , Jingping Liu , Siqi Sun , Kang Liu , Qing-Long Han , and YangTang, "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development" by (2023).DOI: 10.1109/JAS.2023.123618.

[6] Sordoni.A, Galley.M, Auli.M, Brockett.C, Ji.Y , Mitchell.M, NieJ.-Y, Gao.J, and Dolan.B, ''A neural network approach to context-sensitive generation of conversational responses,'' 2015, arXiv:1506.06714.

[7] GuJ.-C, LingZ.-H, and Liu.Q, ''Utterance-to-utterance interactive matching network for multi-turn response selection in retrieval-based chatbots,'' IEEE/ACM Trans. Audio, Speech, Language Process., vol. 28,pp. 369–379, 2020.

[8] Saglam R. B, NurseJ. R. C, and Hodges.D, ''Privacy concerns in chatbot interactions: When to trust and when to worry,'' in HCI International 2021—Posters. Cham, Switzerland: Springer, 2021, pp. 391–399.

[9] Wlla Amayreh and Omar Banimelhem ,"Is ChatGPT a good translator", 2023 14th International Conference on Information and Communication Systems (ICICS),DOI: 10.1109/ICICS60529.2023.10330525. 74

[10]Mostafa M. Amin and Erik Cambria , "Can ChatGPT's Responses Boost Traditional Natural Language Processing", IEEE Intelligent Systems Volume: 38, Issue: 5, Sept.-Oct. 2023),DOI: 10.1109/MIS.2023.3305861