ACPatil
College *of* Engineering

2nd International Conference on
Advances in Technology and Management (ICATM -2024)
April 5-6, 2024

Organized by

A. C. Patil College of Engineering & Management,

Kharghar, Navi Mumbai

www.acpce.org

मानकः पथप्रदर्शकः

INSTITUTION'S INNOVATION COUNCIL
(Ministry of Education Initiative)

ACPCE INNOVATION CELL

ACPatil
College of Engineering

2nd International Conference on
Advances in Technology and Management (ICATM -2024)
April 5-6, 2024

-

PAPER ID: 335

SMARTSPEAK: AI-VOICE ASSISTANT

AUTHOR'S NAME

1. Aniket Varma
2. Vishal Singh
3. Dipanshu Bandoliya
4. Mrs. Rupali Pashte

Outline

Introduction

Literature Review

Problem Definition

Methodology

Results and Discussions

SWOT Analysis

Conclusion

References

Introduction

SMARTSPEAK :

A voice chat bot built using the ChatGPT API and DALL-E API is a web application that allows users to interact with a chatbot using voice commands.

The ChatGPT API provides the chatbot with the ability to understand and respond to natural language, while the DALL-E API provides the chatbot with the ability to generate images.

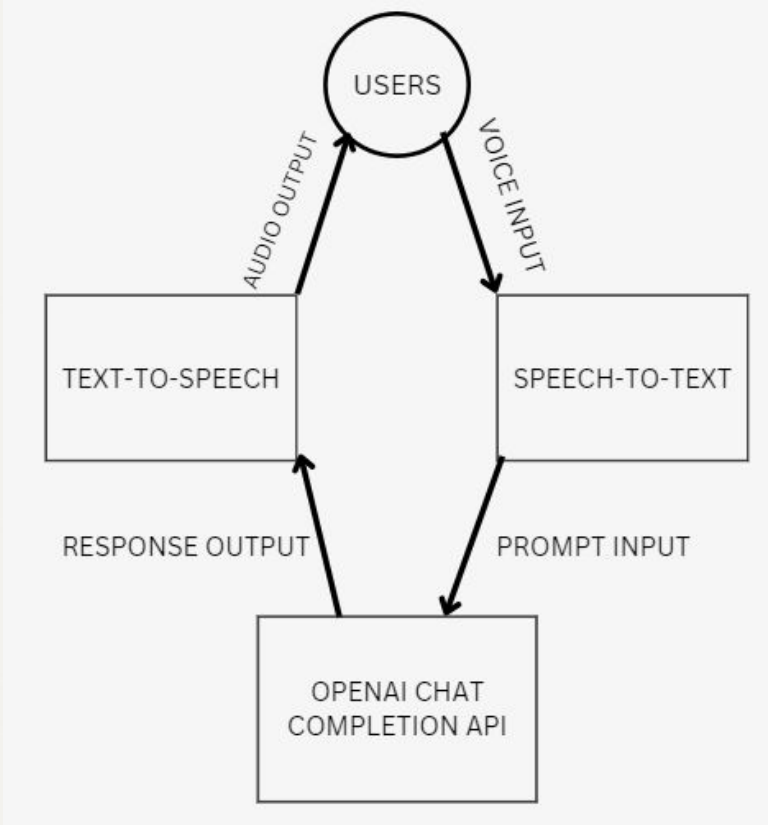
Literature Review

- Daniel Zhang and Jack Rae (2022) introduced ChatGPT, a conversational AI system trained through self-supervision and reinforcement learning to be helpful, harmless, and honest, showcasing remarkable reasoning and interaction capabilities.
- Early work on open-domain chatbots dates back to the 1990s, with recent advancements in large language models like Meena and Blender leading to more sophisticated conversational agents.
- Microsoft's Xiaoice Team (2020) presented an empathetic conversational agent trained on human conversations using inverse reinforcement learning, incorporating external knowledge sources and providing emotional support.

Problem Definition

- The **integration** of voice assistants into daily interactions poses challenges related to limitations in understanding complex queries and providing contextual responses, highlighting the need for enhanced natural language processing and image synthesis capabilities.
- **Privacy and ethical considerations** surrounding data handling and user information security present ongoing challenges in the development and deployment of voice assistant technologies, necessitating robust measures for data anonymization and secure data management.
- **Continuous advancements in voice assistant models** and regular security assessments are essential to enhance reliability and effectiveness, addressing potential vulnerabilities and ensuring the system's adaptability across diverse scenarios and user inputs.

Methodology

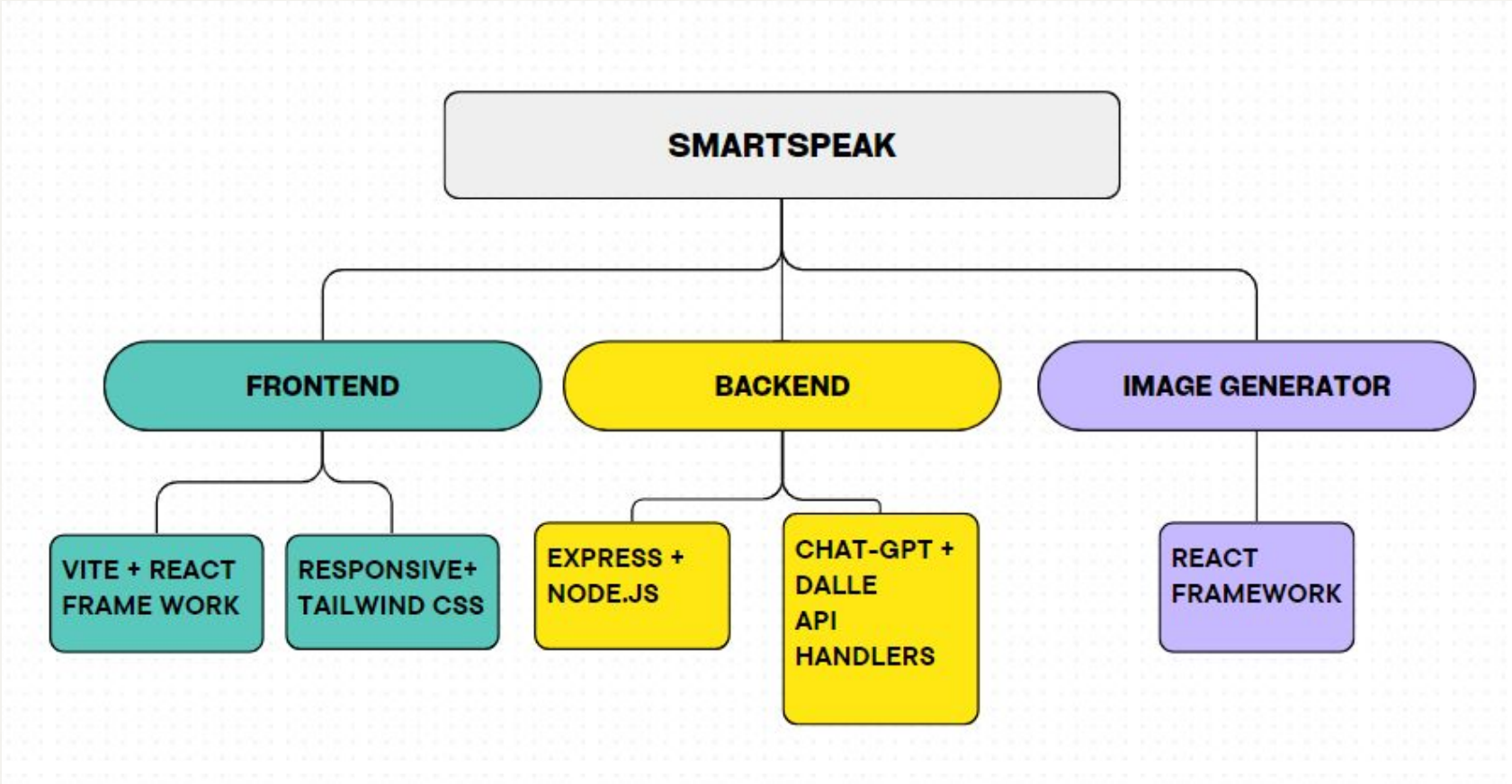


BASIC WORKING OF THE SMART SPEAK WEB APP

RESPONSIVE UI : Works both on mobile and desktop

Latest Technology Used: VITE

Methodology



Project Structure and Technology STACK Used

Results & Discussion

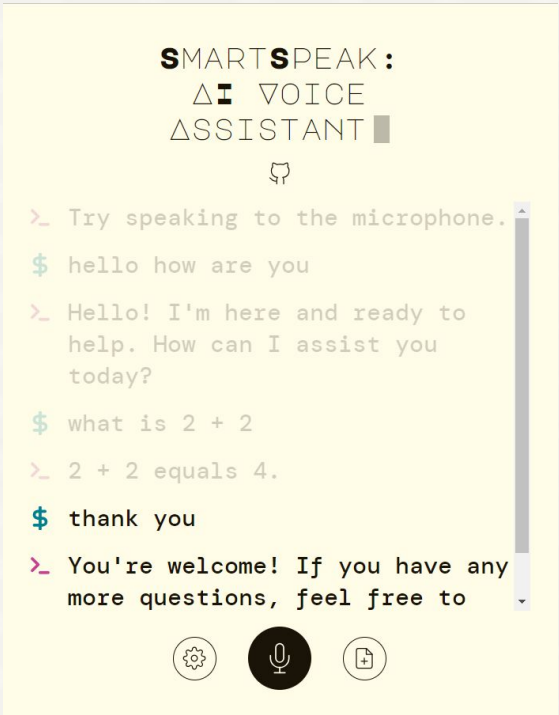


Fig 1. Chatting with SmartSpeak

Chatting with SmartSpeak:

1. The user Clicks the Record button.
2. The user then asks the query through microphone or any other connected devices to the bot.
3. The chat-bot accepts the query and then processes this query in the backend to generate a suitable response.
4. The chat-bot then give the answer as response in the audio format.
5. The user can choose different voice models according to their preferences.

Results & Discussion

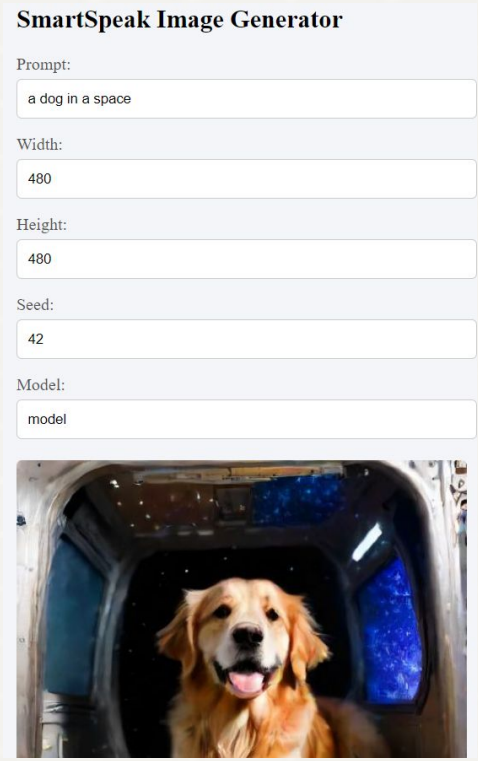
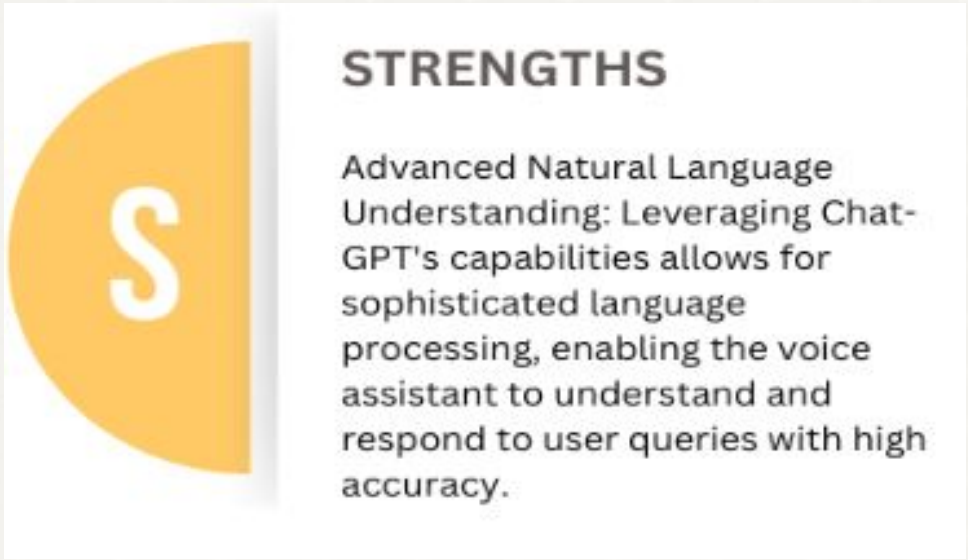


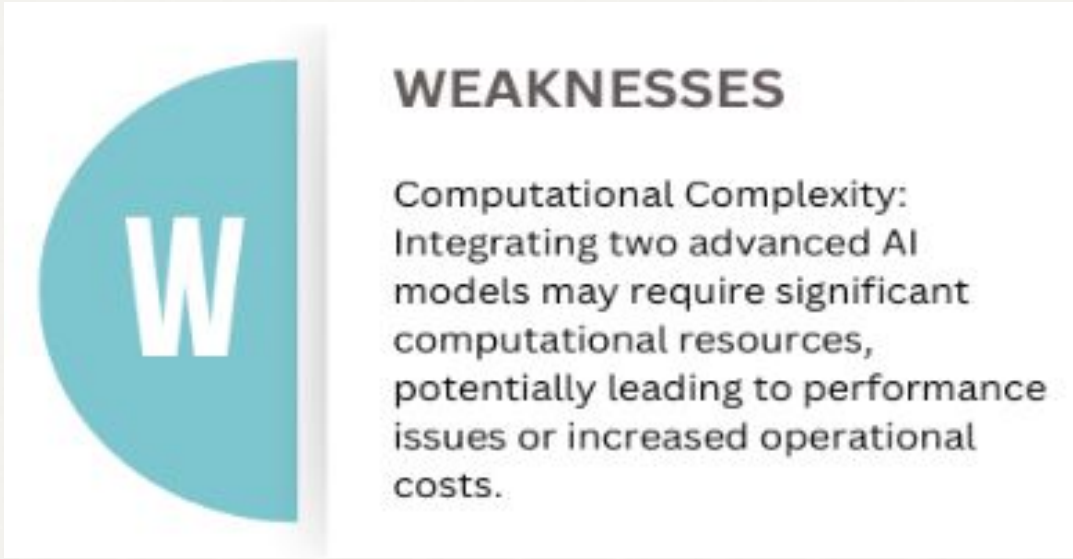
Fig 2. Generating Image with SmartSpeak

Image Generation with SmartSpeak:

1. The user Enters the prompt of the required image.
2. The user then enters the parameters such as height,width of the required image.
3. The chat-bot accepts the prompt and then processes this prompt in the backend to generate a suitable response.
4. The chat-bot then give the response in the image format i.e in jpeg,png etc

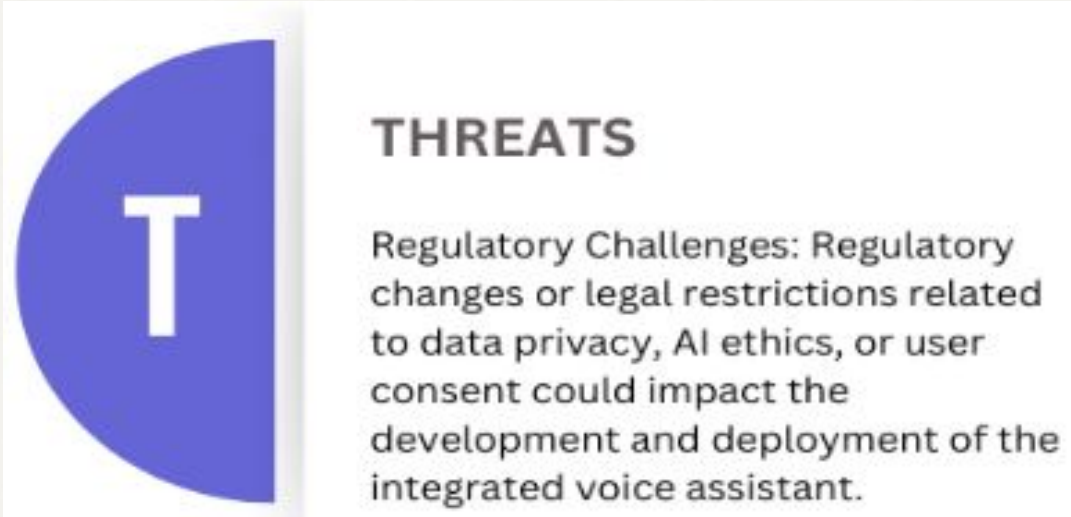
SWOT Analysis





-

SWOT Analysis



Conclusion

In Conclusion,our new system blends cutting-edge technology to talk and understand better than typical voice assistants. It's like having a conversation with a friend who can also show you unimaginable pictures of what you're talking about. This makes it more reliable and user-friendly, giving you a smoother and more enjoyable experience.

References

- Hao, Y., Song, H., Dong, L., Huang, S., Chi, Z., Wang, W., Ma, S., & Wei, F. (2020). Language Models are General-Purpose Interfaces.
- Singh Sikarwar, S. (n.d.). AI BASED VOICE ASSISTANT. Fully Refereed International Journal @International Research Journal of Modernization in Engineering.
- Skantze, G., & Seza Dogruöz, A. (2023). The Open-domain Paradox for Chatbots: Common Ground as the Basis for Human-like Dialogue.
- Zhou, L., Gao, J., Li, D., & Shum, H.-Y. (2020). The Design and Implementation of XiaoIce, an Empathetic Social Chatbot. Computational Linguistics.

Thank You