e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

AI POCKET COMPANION DEVICE

Sunita Joshi[*1], Nidhi Pathak[*2], Dhanshri Londhe[*3], Aditi Satav[*4]

[*1]Lecturer, Department Of Information Technology, Pimpri Chinchwad Polytechnic, Pune, Maharashtra, India.

[*2,3,4]Student, Department Of Information Technology, Pimpri Chinchwad Polytechnic, Pune, Maharashtra, India.

DOI : https://www.doi.org/10.56726/IRJMETS63334

ABSTRACT

This paper presents the development and implementation of an AI-powered Image and Speech Recognition Device using the GPT-4 model, designed to provide users with a seamless, interactive experience. The device is built on a Raspberry Pi 4 microcontroller and equipped with a camera, microphone, speaker, and touchscreen display. Users can interact via typed or spoken commands, and the device processes these inputs through GPT-4 to deliver real-time, accurate responses on the display or speaker. The project explores the application of GPT-4 for handling visual and auditory data, allowing it to function as both an image processing system and a virtual assistant. This device can provide live updates such as weather, news, sports scores, and election results, making it versatile and useful in assistive technology and daily tasks. Testing confirmed the system's efficiency in response accuracy and processing speed, demonstrating its potential for practical, real-world applications. The findings highlight the utility of combining image and speech recognition in a single, portable AI-powered device, offering users a reliable and interactive tool.

Keywords: AI-Powered Device, Image Recognition, Speech Recognition, GPT-4, Real-Time Assistant.

I. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has unlocked new opportunities for integrating smart systems into everyday devices. This project aims to develop a pocket-friendly, AI-powered Image and Speech Recognition Device using a Raspberry Pi 4 microcontroller with the GPT-4 API for processing visual and auditory data. Equipped with a camera, microphone, speaker, and display, the device allows interaction through typed or spoken commands. It performs real-time responses, detecting objects through the camera and processing voice commands via speech recognition, making it suitable for applications like image processing and assistive technology. In addition to these core features, the device functions as a personal assistant by providing daily information such as news, weather, live sports scores, and election results. By combining these capabilities, the device becomes a versatile tool adaptable for a range of practical uses. Leveraging the Raspberry Pi's processing power, the system is efficient, portable, and user-friendly, ideal for enhancing accessibility in various everyday scenarios.

II. METHODOLOGY

The project will be carried out in following phases:

Phase 1: Hardware Setup

In the hardware setup phase, the Raspberry Pi 4 serves as the core microcontroller, integrated with a camera, microphone, speaker, and display screen. The camera is configured for object detection, while the microphone and speaker facilitate voice input and audio output. Each component is tested to ensure compatibility and functionality within the system.

Phase 2: Software Integration

For the software integration, the GPT-4 API is implemented to handle both image recognition and speech processing tasks. The code is developed in Python, with separate modules for image processing, voice recognition, and user interaction. The API key is embedded in the code, allowing the device to access GPT-4's capabilities for interpreting visual and audio inputs.

Phase 3: User Interface Development

The user interface is designed to provide a seamless experience, allowing users to enter commands through text or voice. The display screen presents results and responses generated by GPT-4 in real time. The interface is optimized for simplicity, ensuring that interactions remain intuitive and user-friendly.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

Phase 4: Testing and Validation

During testing and validation, the device undergoes rigorous performance assessments to evaluate its responsiveness, accuracy, and reliability. Image recognition and speech processing functions are tested under different conditions to ensure robustness. Any identified issues are addressed and resolved to enhance system performance.

Phase 5: Deployment and Evaluation

After testing, the device is deployed in a simulated environment to assess its practical utility. Real-time feedback is gathered to make further improvements. Evaluation focuses on the device's versatility in various applications, such as accessing news, weather updates, and real-time election results.

III. MODELING AND ANALYSIS

The device's architecture is modeled with three main layers:

1. **Input Layer:** Consists of a camera and microphone for receiving visual and auditory inputs, allowing users to interact through both typed and spoken commands.
2. **Processing Layer:** The Raspberry Pi 4, integrated with the GPT-4 API, processes inputs. This layer is responsible for interpreting commands, performing image and speech recognition, and generating responses based on the GPT-4 model.
3. **Output Layer:** Comprises a display screen and speaker that provide real-time feedback and responses. The display shows text-based outputs and images, while the speaker vocalizes responses to user commands.

The following table outlines the core components, their specifications, and their roles within the model:

Table 1. Components used in the Pocket Companion

Components	Specification	Functions
Raspberry Pi 4	4GB RAM, Quad-core Cortex-A72 CPU	Acts as the primary processing unit for all tasks
Camera Module	1080p resolution, Raspberry Pi-compatible	Captures images for object detection
Microphone	Super Mini Usb2.0 Microphone	Enables voice input for speech recognition tasks.
Speaker	USB or 3.5mm jack, 5W output	Outputs responses to user commands.
Display Screen	7-inch HDMI touchscreen	Displays text responses and allows user interaction.
GPT-4 API	API key embedded in code	Performs image recognition and processes speech input

The device's performance is evaluated based on three main criteria: accuracy, response time, and user experience.

Accuracy: The device's image recognition and speech processing capabilities were tested across a range of environments and commands. Results indicated high accuracy in object detection and command recognition, reinforcing GPT-4's adaptability to varied input data. The device consistently returned correct responses for standard image and voice inputs, verifying its suitability for practical use.

Response Time: The system's response time was measured by analyzing the time from user input (typed or spoken) to the device's output on the display or speaker. Average response times were recorded between 2-3 seconds, demonstrating the efficiency of the Raspberry Pi 4 and GPT-4 API integration. This fast response rate enhances usability, particularly for applications requiring immediate feedback.

User Experience: User interaction was a core focus in the design, providing a straightforward interface where users can access daily information like news, weather, and sports scores. The device performed well in simulated real-world scenarios, delivering smooth and user-friendly interactions. Feedback from test users indicated satisfaction with the clarity and responsiveness of the commands and responses.

IV. RESULTS AND DISCUSSION

The AI-powered Image and Speech Recognition Device has demonstrated excellent performance in real-time image and speech processing, showcasing its potential as a versatile tool for various applications. This section outlines the successful outcomes from testing, focusing on the effectiveness of the device in image recognition, voice command processing, and real-time information access.

1. **Image Recognition Performance:** The device's image recognition capabilities, powered by the GPT-4 API, proved highly accurate during testing. It successfully identified a wide range of objects captured by the camera under diverse lighting conditions, maintaining high accuracy across tests. This accuracy highlights the device's potential for tasks requiring reliable object detection, making it suitable for applications such as assistive technology and automated systems.
2. **Speech Recognition and Response:** The speech recognition module efficiently processed voice commands, providing accurate and immediate responses through both the display and speaker. GPT-4's language processing enabled the device to interpret a variety of user commands, ensuring that responses remained consistent and reliable across multiple interactions. This functionality supports its role as an interactive assistant, adding value in scenarios where hands-free operation is advantageous.
3. **Real-Time Information Access:** One of the device's standout features is its ability to provide live updates on news, weather, sports scores, and election results. Testing revealed fast response times, averaging between 2-3 seconds, which is ideal for real-time applications. This capability positions the device as a comprehensive source of up-to-date information, enhancing its practicality for users seeking instant access to daily information.

Discussion

The results affirm the effectiveness of combining GPT-4's advanced language and image processing with the Raspberry Pi's capabilities, resulting in a compact and powerful device. This AI-powered system can seamlessly support applications across various domains, offering users a reliable, interactive experience that adapts to diverse needs. With its successful performance in real-time image recognition, voice command processing, and live information retrieval, the device presents a strong foundation for future AI-powered applications in assistive and personal technology.

V. CONCLUSION

This project developed a versatile, AI-powered device integrating image and speech recognition using a Raspberry Pi 4 and GPT-4 API. The device performs real-time processing for both visual and auditory commands, delivering fast, accurate responses. With its additiossnal functionality as a personal assistant, the device provides users with access to information such as news, weather, sports, and election results, enhancing its utility beyond standard recognition tasks. The successful integration of these features in a compact, accessible format highlights the potential of combining advanced AI with affordable microcontroller technology. This device represents a practical solution for diverse applications, offering a strong foundation for further innovation in AI-driven, interactive systems.

VI. REFERENCES

[1] GPT-4 Technical Documentation OpenAI. GPT-4 Technical Report.

[2] Kaymak, C., & Ucar, A. (2018). Implementation of Object Detection and Recognition Algorithms on a Robotic Arm Platform Using Raspberry Pi. Proceedings of the 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 1-8. DOI:

[3] Lee, R., et al. "AI in Assistive Technologies: Applications and Real-time Challenges." International Journal of Assistive Technology, 2020

[4] Brown, A., et al. "Implementing Real-Time AI Solutions on Edge Devices: A Study with Raspberry Pi and GPT APIs." Journal of Artificial Intelligence Research and Development, vol. 35, no. 4, 2022, pp. 210-225.

[5] Chen, L., et al. "Deploying AI-Powered Image and Speech Recognition on Edge Devices: Applications with Raspberry Pi and Cloud Integration." International Journal of Applied Artificial Intelligence, vol. 30, no. 2, 2023, pp. 145-160