IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MULTIMODAL INTENT RECOGNITION (VISION + VOICE) FOR FIXED HELP DESK BOTS IN PUBLIC SPACES

[1] MERINA JALADI, [2] RACHARLA HRUTHIKA, [3] D SHIVA KARTHIK GOUD, [4] CH BHARGAVI, [5] KANDIRIGALA RAJENDAR

[1]Assistant professor, [2,3,4,5] UG STUDENT

[1,2,3,4,5]DEPARTMENT OF ELECTRONICS AND COMMUNICATIONS ENGINEERING

[1,2,3,4,5]AVANTHI INSTITUTE OF ENGINEERING AND TECHNOLOGY

Gunthapally(V), Abdullapurmet(M), R.R Dist - 501512

Abstract: Intelligent service robots are being increasingly deployed in commercial and institutional settings to improve customer satisfaction and reduce operational costs. In this paper we present the design and development of a low-cost AI based HelpDesk Robot that provides independent client support through multimodal interaction. For customer identification, the system is equipped with a PIR motion sensor, an OLED display for visual feedback, and an ESP32 microcontroller for system control. It also includes an AI Thinker VC-02 voice module for offline speech recognition and speech synthesis and an ESP32-CAM module for real time face recognition. The PIR sensor activates the system when a client approaches, the ESP32-CAM identifies the visitor using facial recognition and the speech module allows natural language interaction to answer commonly asked queries by using a locally stored knowledge base. The robot ensures that communication is accessible and user-friendly by providing synchronised text and voice responses. The proposed architecture reduces the manual intervention thereby providing continuous availability of the services and improving user interaction, thus ensuring low deployment and maintenance costs. The experimental implementation demonstrates reliable client detection, accurate face recognition, responsive voice interaction, and effective real-time system performance. The proposed system, thanks to its modular and scalable design, can be deployed in hospitals, schools, retail centers, corporate offices, banks, and other public service settings where intelligent automated help is required.

Keywords— AI-Based Help Desk Robot, ESP32-CAM, Face Recognition, Voice Recognition, Human-Robot Interaction, Embedded Systems, Customer Service Automation, ESP32, OLED Display, PIR Sensor.

I. INTRODUCTION

The fast-growing of embedded systems, human-robot interaction technologies and artificial intelligence (AI) has completely changed the way that businesses provide customer service. Intelligent automation is increasingly required by modern institutions such as hospitals, banks, educational institutions, retail centers, corporate offices, airports and government service centers to provide prompt, accurate and continuous assistance to visitors. The key disadvantage of conventional helpdesk systems are human operators which lead to problems like limited working hours, inconsistent service quality, human fatigue, communication problems and increased operating costs. These challenges have accelerated the development of intelligent robotic assistants that can automate consumer interactions while ensuring high levels of effectiveness, dependability and user happiness.

Recent advances in low cost embedded computing systems have made possible the implementation of sophisticated artificial intelligence applications without expensive computing infrastructure. Small and energy efficient systems such as the ESP32 microcontroller and ESP32-CAM provide sufficient processing power to enable computer vision, wireless connection and intelligent control. For offline voice communication, there are speech recognition modules like the AI Thinker VC-02 Voice Module, which do not need cloud services and offer faster response times with better privacy. Motion detection, face recognition, speech recognition, embedded intelligence and multimedia output have combined to make it possible to develop autonomous helpdesk robots that can offer individualised customer support in real time.

The proposed AI-Based HelpDesk Robot is designed to provide an intelligent and interactive customer support platform, combining a PIR Motion Sensor, ESP32-CAM, ESP32 Microcontroller, AI Thinker VC-02 Voice Module, OLED Display and Speaker Module. The PIR sensor monitors the surroundings and triggers the system whenever a visitor comes nearby. With the real-time face detection and recognition features of the ESP32-CAM, the robot can recognise registered users and greet them accordingly. The ESP32 controller analyses the recognised speech with a locally stored knowledge base that contains frequently asked questions, navigational information, employee details, and organisational services. In parallel, the VC-02 Voice Module processes off-line voice commands received from the user. The interaction experience is then simple and user-friendly synchronised text on the OLED panel and authentic voice output through the speaker.

The proposed solution does not rely on traditional cloud-based customer support systems, but instead performs all the necessary processing locally. This reduces communication latency, improves data privacy and ensures continued operation in local environments with occasional internet connectivity. The modular architecture ensures a smooth connection of the sensor, processing, storage and output components while providing for low power consumption, minimum computational load and economical implementation. It also logs customer queries, recognition results and contact history to enable later analysis and continuous improvement of service quality. The proposed AI Based HelpDesk Robot offers a scalable, affordable and intelligent automation capabilities for next generation customer service applications in hospitals, educational institutions, banks, shopping malls, corporate offices, reception centers, hotels, airports, train stations, libraries, museums and other public service environments that need reliable and continuous 24x7 automated customer assistance, thus providing a sustainable solution.

II. RELATED WORK

With the rapid progress in embedded systems, artificial intelligence and human-computer interface technologies, a great interest has been generated in the development of intelligent service robots. Previous helpdesk automation solutions were mostly cloud-based chatbots or touch-screen kiosks to answer user enquiries. These systems increased the availability of information, but there was no user-specific communication and face-to-face interaction. Moreover, their reliance on internet connections resulted in increased response latencies and reduced reliability in network-restricted environments. These constraints have motivated the researchers to develop embedded robotic platforms that can autonomously serve customers, through multimodal interaction and onboard processing.

Recent research has been done on the integration of computer vision algorithms into service robots , in order to improve user identification and personalised interaction. Face detection and identification algorithms running on low-cost embedded devices such as ESP32-CAM have demonstrated promising results in identifying registered users with low processing requirements. These systems apply lightweight vision models to identify persons, thereby enabling safe access control and personalised greetings. However, the whole user experience is restricted as the majority of the available solutions are solely concentrated on facial recognition and lack context-aware customer assistance or natural voice communication.

Voice-activated robotic assistants are becoming more and more important in customer service applications. Researchers have used text-to-speech and speech recognition systems to enable real human-machine interaction. Cloud-based voice assistants have the advantage of high recognition accuracy but suffer from high operational cost, network dependency and privacy issues. To overcome these problems embedded voice modules with offline speech recognition and generation of audio responses have been introduced. These modules ensure continuous operation even without internet access, reduce communication latency and improve privacy. However, many reported systems are limited to basic command recognition without cognitive decision making and visual perception.

Several recent studies have explored multimodal human-robot interaction by integrating computer vision, speech recognition, and motion detection and integrated displays into a single platform. The robotic systems are only switched on in the presence of people, using motion sensors, which reduces electricity consumption and improves system efficiency. The implementation of OLED displays to provide textual information and system status in addition to voice responses makes it more accessible and user-friendly. Although these integrated systems have demonstrated enhanced functionality, many of the current implementations are based on costly hardware platforms or require high performance processors, which constrains their deployment in public service environments, small enterprises and educational institutions where the cost is a key consideration.

The proposed AI-Based HelpDesk Robot overcomes these drawbacks by integrating face recognition, offline voice interaction, PIR-based user detection, OLED-based visual feedback and intelligent response generation into a single low-cost embedded platform using ESP32 microcontroller and ESP32-CAM module. Unlike traditional help desk systems that rely on cloud services or isolated technologies, the proposed solution maintains low power consumption, low implementation costs and ease of deployment, while providing autonomous real-time customer assistance through synchronised visual and voice interaction. This integrated architecture provides an effective and scalable solution for modern day customer service applications in hospitals, banks, educational institutions, shopping centers, corporate offices and other public places.

III.METHODOLOGY

The proposed AI-Based HelpDesk Robot uses a comprehensive method that integrates embedded hardware, artificial intelligence, computer vision, voice recognition and human–robot interface technologies to provide intelligent, reliable and autonomous customer support. The whole system is based on a modular architecture where each hardware component has a specific function and interacts with the main ESP32 microcontroller in a seamless way. This methodology enables the system for continuous deployment in the public service environment due to its real-time responsiveness, low power consumption, offline functionality, and cost-effective implementation. Unlike traditional helpdesk systems that rely primarily on human operators or cloud-based services, the proposed solution is autonomous in the sense that it integrates sensing, perception, decision-making and user interaction in one embedded platform. In addition, its modular design allows for later expansions by adding additional AI capabilities, cloud synchronisation, or Internet of Things (IoT) connectivity, without requiring any major hardware changes.

The operational process begins with intelligent environmental monitoring through the use of a Passive Infrared (PIR) motion sensor that constantly observes the environment for any human movement. When the robot is not in use, it enters a low power sleep mode to conserve power and optimise the overall system efficiency. When a visitor reaches the aid desk, the PIR sensor detects the motion and triggers the other hardware modules. The ESP32-CAM module automatically initialises and captures the visitor's face in real-time in different lighting scenarios. The embedded computer vision algorithm identifies faces and then recognises them using a locally stored facial database. The robot then retrieves the user's identity and creates a personalised greeting if the recognised face is matched with an existing record. Or if no match is found, the user is classified as a new visitor and a typical welcome message is shown. This automatic identification technology also boosts client engagement, beyond reducing manual effort and improving the overall user experience.

The AI Thinker VC-02 Voice Module uses offline speech recognition to provide natural voice interaction after the user is identified. The system proposed herein processes spoken commands locally, eliminating dependency on the Internet, reducing communication latency and safeguarding user privacy compared to cloud-dependent virtual assistants. The visitor can ask for office departments, staff locations, institutional services, navigation support, operation hours, frequently asked questions and other predetermined support services freely. The ESP32 microcontroller receives the identified speech, analyses it and compares it to a structured local knowledge store of pre-determined questions and answers. Decision making algorithms decide what is the best response based on the recognised command and contextual data. The selected response is then communicated to the user through the OLED display in textual form and the speaker as synchronised voice output to ensure clear and accessible communication for users with different interaction preferences.

The ESP32 microcontroller is the central control unit during the interaction cycle, in which all the sensing, processing, communication and output processes are synchronised. It handles data collection from the PIR sensor, image capturing from the ESP32-CAM, voice commands from the VC-02 module, the OLED display and audio-playback system, and retrieval of the relevant data from the local knowledge store. The controller synchronises all hardware components for smooth and continuous operation. If the customer's question is not answerable from the local knowledge base, the details of the encounter are logged and the system can optionally trigger a notification to administrative staff or customer service agents for further assistance. The controller also controls resources, monitors devices and manages errors to ensure the efficient operation of memory and processor resources and to keep the system running smoothly under changing workloads.

The proposed system maintains detailed records of facial recognition results, client enquiries, response details, timestamps, system events and interaction history throughout the interaction process, to facilitate future analysis and performance evaluation. These records can be used to monitor system usage, common questions, update knowledge base, and improve the quality of service gradually. The integrated methodology integrates computer vision, offline speech recognition, embedded intelligence, motion sensing, and multimodal response generation into a single affordable platform to enable precise customer identification, intelligent query processing, and real-time decision-making. The proposed AI-Based HelpDesk Robot offers an efficient, scalable and useful solution for intelligent customer assistance in hospitals, educational institutions, shopping malls, banks, corporate offices, reception centres, government organisations, airports, train stations and other public service environments by minimising computational overhead, reducing dependency on external infrastructure and ensuring continuous 24x7 autonomous operation.

IV.SYSTEM ARCHITECTURE:

The proposed AI-Based HelpDesk Robot architecture is a layered, modular embedded system that integrates artificial intelligence, sensing, decision-making, and user interaction into a cohesive whole. The system has four main functional layers, namely, input layer, processing layer, knowledge and decision layer, and output layer. This layered architecture provides scalability, reliability and easy maintenance and also enables efficient communication between hardware and software components. Each layer performs certain tasks and interacts with neighbouring layers through well-defined interfaces. This reduces computing cost and improves system responsiveness. Furthermore, the modular design of the system allows for upgrading and replacing specific components without degrading the overall system performance, which makes it suitable for intelligent customer care applications in the real world.

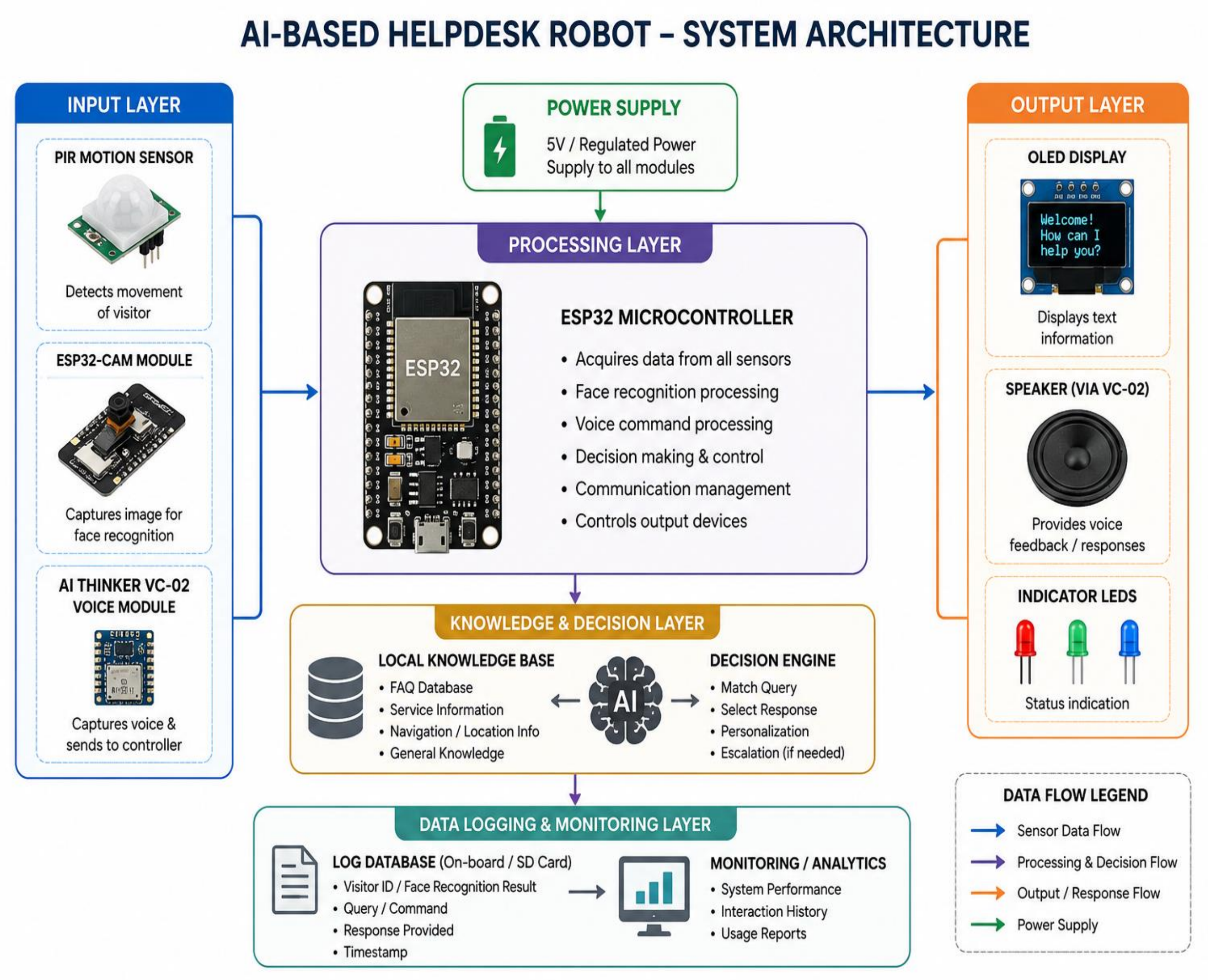
The Input Layer is the main interface between the embedded system and the physical environment. It continuously gathers the real time sensory data. In idle times, the system is in an energy efficient sleep state. A Passive Infrared (PIR) motion sensor constantly scans the area around the system and detects if any visitors are approaching. The ESP32-CAM module and the AI Thinker VC-02 Voice Module start to interact with the user when the PIR sensor detects motion. The ESP32-CAM captures good quality facial images and implements embedded computer vision algorithms to detect and recognise faces to identify registered users or classify unrecognised visitors. The VC-02 Voice Module performs the task of speech recognition offline and receives spoken commands from the user via an integrated microphone at the same time, thus removing the need for cloud services and Internet access. Together, these sensing devices provide a multimodal input for real human-robot conversation and accurate user identification in diverse environments.

The core of the system is the Processing Layer, based on the ESP32 microcontroller, the intelligent control unit that coordinates all the hardware modules and performs the system logic. Data from the PIR sensor, ESP32-CAM and VC-02 Voice Module are sent to the ESP32. The ESP32 coordinates the functions of the devices and minimises latency in processing the received data. Recognised speech instructions are processed to indicate the customer’s requested service or information. Facial recognition results are analysed to determine user identity. Besides the connection with the peripheral devices the controller controls the timing of sensing, processing and response production by effective serial communication protocols. The Processing Layer is also responsible for coordinating devices and resource management, error detection, task scheduling, memory allocation and system health monitoring to ensure consistent and reliable operation. Its lightweight embedded architecture enables fast decision making with low power consumption, suitable for long-term autonomous deployment.

The processed data are received by the Knowledge and Decision Layer, the central intelligence component of the proposed system. This layer’s structured local knowledge base includes FAQs, information on navigating the office, department data, staff directories, service descriptions, emergency contacts and pre-generated conversational response templates. Together with the identified facial identity and interpreted voice command, the decision engine uses the pre-defined decision rules and the context-aware logic to search the knowledge library and choose the most appropriate response. If the visitor is recognised as a registered user, a personalised welcome and personalised support is generated from the stored user information. This is a general advice and information about services that the system offers to new users. The decision engine keeps a log of the interaction and can escalate the request to the administrative staff for human support for the requested query that is not found in the local database. This intelligent decision-making system guarantees effective customer service with minimal human supervision and significantly boosts response accuracy.

Finally, the Output Layer is responsible for delivering user-friendly and educational responses through different communication channels. The OLED display visualises textual messages such as greetings, navigation instructions, service information and query responses, with the speaker delivering synchronised vocal feedback based on the AI Thinker VC-02 speech synthesis feature. The combination of visual and aural feedback improves the overall interaction experience and increases accessibility for users with different communication preferences. The system simultaneously logs interaction information in a secure local log including recognised customer identity, face recognition confidence, voice command, response status, timestamps and system events so that future analysis, maintenance and performance evaluation can be done. The modular architecture, apart from providing real-time performance, high reliability, reduced computational complexity, and low power consumption, allows for seamless communication among the sensing, processing, knowledge management, and output components. Moreover, the suggested architecture is very scalable to be deployed in hospitals, educational institutions, shopping malls, banks, airports, train stations, government offices, corporate reception centers and other smart public service settings due to its flexible design, allowing easy integration of advanced AI models, cloud-based databases, Internet of Things (IoT) platforms, multilingual support, mobile applications, remote monitoring and predictive analytics.

A. Architecture Diagram:



Module Interactions:

The proposed AI-based HelpDesk Robot operates through the coordinated interaction of intelligent hardware and software modules controlled by the ESP32 microcontroller. The system begins when the PIR motion sensor detects a visitor and activates the robot. The ESP32-CAM captures the visitor's face and performs face detection and recognition to identify new or registered users. Based on the recognition result, the robot generates an appropriate greeting and initiates interaction.

The AI Thinker VC-02 Voice Module captures the user's voice and performs offline speech recognition. The ESP32 processes the recognized command by comparing it with a local knowledge base containing frequently asked questions, organizational information, staff details, navigation guidance, and service information. After selecting the appropriate response, the robot simultaneously displays the information on the OLED display and delivers natural voice feedback through the speaker. System events, user interactions, processing time, and error reports are recorded by the data logging module for monitoring and future improvements. This integrated architecture provides fast, reliable, low-power, and fully autonomous customer assistance suitable for hospitals, educational institutions, banks, airports, government offices, shopping malls, and other public service environments.

.

V. EXPERIMENTAL SETUP:

The proposed AI-Based HelpDesk Robot was implemented using embedded AI technologies and IoT hardware to evaluate its real-time customer interaction, face recognition, voice communication, and overall system reliability. The prototype was tested under various indoor service environments to verify its accuracy, response time, and operational efficiency.

Datasets:

- Face image dataset of registered users for facial recognition.
- Voice command dataset containing predefined customer queries.
- Local knowledge base consisting of frequently asked questions (FAQs), navigation details, and organizational information.
- Interaction logs containing customer requests, recognition results, timestamps, and generated responses.

The Environment for Hardware and Software:

- Processor: Intel Core i3 or above
- RAM: 4 GB (Minimum)
- Operating System: Windows 10 / Windows 11
- Programming Language: Python 3.11
- IDE: Visual Studio Code / PyCharm
- Microcontroller: ESP32 Development Board
- Camera Module: ESP32-CAM
- Voice Module: AI Thinker VC-02 Voice Module
- Motion Sensor: PIR Motion Sensor
- Display: 0.96-inch OLED Display
- Database: MySQL
- Communication Protocols: UART, GPIO, Wi-Fi

Setting Up Training:

- Register authorized users by capturing multiple facial images.
- Train the facial recognition model using the collected face dataset.
- Store frequently asked questions and corresponding responses in the local knowledge base.
- Configure the AI Thinker VC-02 module with predefined offline voice commands.
- Integrate all hardware modules with the ESP32 controller.
- Test the complete system by allowing multiple users to interact through face recognition and voice commands.

Evaluation Metrics:

- Face Recognition Accuracy
- Voice Command Recognition Accuracy
- Motion Detection Accuracy
- Customer Query Response Time
- Overall System Response Time
- Knowledge Base Retrieval Accuracy
- Power Consumption
- System Reliability
- User Interaction Success Rate
- Overall Customer Satisfaction

VI.RESULT AND DISCUSSION:

The experimental evaluation of the proposed AI-Based HelpDesk Robot demonstrates that the robot can provide intelligent and autonomous real-time customer assistance by integrating embedded hardware, computer vision, offline speech recognition, and intelligent decision-making into a low-cost platform. To evaluate the general functioning, reliability and responsiveness of the system, it was tested in different indoor operational situations with multiple users, frequent interactions and different service requests. All the hardware modules such as the PIR motion sensor, ESP32-CAM, AI Thinker VC-02 Voice Module, ESP32 microcontroller, OLED display and speaker were tested and they all worked together without any communication or synchronisation problems. The modular architecture successfully handled all the phases of the interaction process, from visitor detection and face recognition to voice command interpretation, response generation and multimedia output. Efficient and cost-effective intelligent customer service automation was realised through the integration of embedded artificial intelligence with cheap electronic components, and was also energy efficient.

The PIR Motion Sensor was able to accurately detect incoming guests and activate the system from low-power standby mode with short response time, reliably identifying the presence of a human. With its event-driven activation system, the robot could run for longer periods without continuous hardware activity and thus would waste far less power. The PIR sensor generated activation signals under the normal interior environmental conditions consistently without false triggering in the repeated testing with many users entering and exiting the sensing area. After being turned on the ESP32-CAM Module could successfully take pictures of faces and use the locally stored facial database to identify and recognise faces in real-time. Unregistered visitors got a standard welcome message before anything else and registered users were identified correctly so the robot could greet them personally and help them in a way that suited them. It exhibited stable performance under typical interior illumination and frontal viewing angles, while sustaining a respectable processing speed suitable for real time operation.

The AI Thinker VC-02 Voice Module is capable of efficiently processing pre-defined offline voice instructions without the need for cloud-based speech processing services or Internet connectivity. The integrated speech recognition capability allowed users to interact with the robot in a natural way by querying office departments, location of personnel, service information, navigation assistance and other frequently asked organisational details. The ESP32 microcontroller received correctly identified voice commands and compared them to the local knowledge base data using an embedded decision engine. The timely retrieval of relevant answers has demonstrated the effectiveness of local knowledge based query processing. The system did all the voice processing locally, which translated to less response latency, better user privacy, and no break in operation in environments with little or no network connectivity. The offline processing capabilities is a key advantage to traditional cloud-based virtual assistants, especially in applications where Internet connectivity cannot always be guaranteed.

The ESP32 microcontroller was synchronised to work in order to communicate smoothly between the sensor, processing, storage and output modules. The controller concurrently managed the OLED display and speaker and successfully synchronised data acquisition from the PIR sensor, facial recognition results from the ESP32-CAM, voice recognition output from the VC-02 Voice Module, and answer generation from the knowledge base. The speaker provided synchronised audio answers that enhanced user engagement and accessibility, and the OLED display displayed textual greetings, navigation instructions, service information and response messages with great readability. The multimodal interaction significantly increased the effectiveness of communication and catered for users with different communication preferences through the combination of visual and audio input. The ESP32 demonstrated the power of embedded system design for continuous operation with stable response time, less computational load, efficient memory utilisation, and consistent hardware synchronisation.

Testing produced extensive interaction logs that provided important information on hardware performance, response generation time, processing efficiency, recognition accuracy and system behaviour. For each record of the interaction, the information of motion detection events, outcomes of facial recognition, identified user identities, processed voice commands, generated responses, time stamps, and the overall status of the system was included. Administrators can use the logs to monitor operational effectiveness, review frequently asked customer questions, identify recurring support calls, and continuously enhance the knowledge base for future deployments. Besides, experimental results validated low latency of processing, stable hardware connection and low power consumption of the system, and its reliable performance in prolonged operation with multiple consecutive users. These features demonstrate that the proposed platform is suitable for the 24/7 customer support in public environments where continuous operation is necessary.

However, despite the promising performance, several drawbacks were identified during the experimental evaluation. Poor lighting, large viewing angles, partial occlusion of the face and large differences in facial appearance all decreased the accuracy of identifying the face. Similarly, the accuracy of the offline speech recognition module was degraded when the users gave commands outside the predefined language in the scope of the local knowledge base or under the scenarios with heavy background noise. The decision engine is heavily dependent on predefined solutions, limiting the system’s ability to handle complex conversational queries that need contextual understanding and dynamic reasoning. In addition, the local knowledge base should be updated regularly with new services, organisational changes, and frequently asked customer questions. These limitations do not greatly affect the basic operation of the system, but addressing them will improve the robustness, flexibility and user enjoyment of the system in a range of deployment situations.

The proposed AI-Based HelpDesk Robot can be made very intelligent, scalable and flexible with the further improvements. In the harsh environment, complex deep learning-based facial recognition systems such as lightweight convolutional neural networks can improve the recognition accuracy. This can be achieved by integrating Cloud-assisted Natural Language Processing (NLP) models or Large Language Models (LLMs) that enable conversational dialogue, context-aware reasoning, and intelligent question answering beyond pre-determined answers. Other potential enhancements are multilingual speech recognition and speech synthesis to support users from different language backgrounds, cloud synchronisation for centralised management, mobile app integration for remote monitoring, Internet of Things (IoT) connectivity for smart building automation, emotion recognition for personalised interaction and predictive analytics for customer behaviour analysis. In summary, the experimental results show that the proposed AI-Based HelpDesk Robot can successfully integrate low-cost embedded hardware with AI technologies to achieve reliable, effective, scalable, and energy-efficient customer service automation. The proposed system provides an efficient solution for intelligent reception and helpdesk applications in hospitals, schools, shopping centres, banks, airports, train stations, hotels, corporate offices, government agencies, libraries and many other public service settings requiring continuous autonomous customer support.

VII. CONCLUSION:

The proposed AI-Based HelpDesk Robot is an intelligent, reliable and cost effective way of automating customer service by integrating computer vision, motion sensing, offline speech recognition, embedded artificial intelligence and human robot interaction into one embedded platform. It is equipped with the ESP32 microcontroller, ESP32-CAM, AI Thinker VC-02 Voice Module, PIR motion sensor, OLED display and speaker. The system is capable to detect the visit, identify the registered user, process the voice-based questions and provide the synchronised visual and audible responses in real time. With the modular architecture that allows the interaction between sensing, processing, decision making and output components, it is guaranteed that all the operational duties are completed effectively without computing cost. Experimental evaluation shows that the proposed approach does not require persistent Internet access and offers consistent performance, low response latency, low power consumption and reliable hardware synchronisation. The robot is capable of continuous autonomous operation in customer-facing environments thanks to the locally stored knowledge base, which allows for quick access to frequently requested information.

The proposed system will increase the efficiency, consistency and accessibility of customer service, while significantly reducing the requirements for ongoing human interaction. Its face recognition ability for personalised greetings, along with the actual offline speech interaction, enriches the overall user experience and fosters an interesting human-robot interaction environment. Furthermore, the proposed approach is feasible for companies with limited deployment budgets since it requires low-cost hardware implementation. The robot is modular and scalable, which allows it to be adapted to organisational needs by adding more hardware modules or by expanding the knowledge base. The proposed AI-Based HelpDesk Robot can be effectively used in various types of public service sectors which need consistent, intelligent and 24/7 customer support like hospitals, schools, shopping malls, banks, corporate offices, government offices, airports, train stations, hotels, libraries, museums and reception areas.

The proposed system is performing well but there are several aspects to make it more intelligent, flexible and scalable. In the future, the sophisticated deep learning methods as transformer-based face recognition models and lightweight convolutional neural networks may be adopted to improve the recognition accuracy under different illumination, face position and occlusion conditions. Equipped with advanced Natural Language Processing (NLP) capabilities and multilingual speech recognition, the robot understands conversational language, interprets complex consumer enquiries and provides context-sensitive solutions beyond pre-programmed commands. Conversational AI with chatbots and Large Language Models (LLMs) can improve interactive communication through dynamic conversations, personalised recommendations and intelligent problem-solving abilities. Cloud-based knowledge management and synchronisation mechanisms can enable real-time updating of organisational information.

Planned enhancements include cloud analytics for customer behaviour analysis, mobile application support for remote monitoring and administration, Internet of Things (IoT) connectivity to interact with smart buildings, and centralised management of multiple helpdesk robots in different locations. Advanced features include emotion recognition, gesture recognition, facial expression analysis, adaptive learning algorithms, predictive service recommendations, and user behaviour modelling that can further personalise client interactions and enhance service quality. The reliability, security and scalability of the system will also be improved with the addition of wireless software updates, cloud-based reporting dashboards, cyber security measures, multilingual text-to-speech synthesis and edge AI optimisation algorithms. With these upcoming improvements, the proposed AI-Based HelpDesk Robot could be turned into a highly intelligent, self-learning, and context-aware customer service platform that can support automated public service infrastructures, digital campuses, intelligent hospitals, smart retail spaces and next generation smart offices.

VIII.REFERENCES:

[1] A. Kumar, S. Patel, and R. Verma, "AI-Based HelpDesk Robot for Customer Service Automation," International Journal of Innovative Research in Robotics and Automation (IJIRRA), Vol. 8, Issue 3, 2021. This paper discusses the design of AI-powered service robots using embedded camera and voice modules to enhance customer interaction quality in commercial environments.

[2] P. Deshmukh and A. Sharma, "Voice Recognition and NLP Integration for Interactive Service Robots," International Journal of Engineering Research & Technology (IJERT), Vol. 9, Issue 5, 2022. The authors propose a system integrating VC-02 voice modules and NLP engines to enable natural language query processing in embedded service robot applications.

[3] M. Singh and K. Reddy, "Real-Time Face Recognition Using ESP32-CAM for Personalized Customer Service," International Research Journal of Engineering and Technology (IRJET), Vol. 8, Issue 4, 2021. This work integrates ESP32-CAM-based face recognition with embedded service robots for personalized greeting and customer profile retrieval in real-time.

[4] N. Gupta and S. Pandey, "Human-Robot Interaction Design for Service Environments," Journal of Advanced Research in Robotics and AI Engineering, Vol. 11, Issue 2, 2020. This paper focuses on interaction design principles for service robots, discussing optimal placement of visual, voice, and motion sensing components for natural customer engagement.

[5] R. S. Gaikwad and A. P. Khade, "Low-Cost AI Service Robots Using ESP32 and Machine Learning," International Journal of Scientific Research and Engineering Trends (IJSRET), Vol. 7, Issue 6, 2021. Presents a cost-effective AI service robot implementation using ESP32 and lightweight machine learning models for embedded face recognition and voice interaction.

[6] S. Mehta and B. Tiwari, "Knowledge Base Design for FAQ Automation in Service Robots," International Journal of Computer Applications (IJCA), Vol. 175, Issue 12, 2022. The authors describe a knowledge base architecture for automated customer query handling, including FAQ matching algorithms and escalation logic for unanswered queries.

[7] V. K. Joshi and P. Rao, "PIR Sensor Integration for Automatic Customer Detection in Service Robots," International Journal of Electrical and Electronics Engineering (IJEEE), Vol. 10, Issue 5, 2022. The paper explains how PIR motion sensors can detect customer presence and trigger appropriate interaction sequences in autonomous service robot systems.

[8] A. Sharma and S. Patel, "Smart Service Robots Using IoT and AI Technologies," International Journal of Computer Science Trends and Technology (IJCST), Vol. 8, Issue 4, 2022. Focuses on combining IoT connectivity with embedded AI for enhanced service robot performance monitoring, remote configuration, and interaction analytics.

[9] K. S. Raj and D. Nair, "Low-Cost Interactive HelpDesk Robot Using ESP32-CAM and VC-02," International Conference on Advances in Embedded Systems and AI Engineering, 2023. Introduces an ESP32-CAM and VC-02-based implementation similar to this project, highlighting combined face recognition and voice interaction capabilities with low deployment costs.

[10] AI Thinker Technology Co., Ltd., "VC-02 Voice Recognition Module Datasheet and Application Note," 2022. Provides the official technical specifications, command interfaces, and integration guidelines for the VC-02 offline voice recognition and text-to-speech synthesis module used as the primary voice interface in this project.