



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 5, May 2026

Robex A Next Generation AI Powered Robotic System

Prof. Shegar S.R, Late Durga Mahadeo

Assistant Professor, Department of Computer Engineering, Samarth College of Engineering, Belhe, India

Department of Computer Engineering, Samarth College of Engineering, Belhe, India

ABSTRACT: Artificial Intelligence and the Internet of Things have really improved a lot lately. This means we can now build systems that can work in time know what is happening around them and do tasks on their own. I want to show you a platform that uses Artificial Intelligence for schools. This Artificial Intelligence platform is really cool because it is a robot that can move around by itself and talk to people with a voice. The Artificial Intelligence robot can even recognize faces to take attendance. The Artificial Intelligence robot works because of a design that it uses. The system has a parts: a hardware part, a processing part and a cloud data part. I made the hardware part using sensors and motors. I wrote code for the processing part so it can use intelligence for recognizing voices and faces. I linked the processing part to the cloud data part to keep track of attendance. When I tried out the intelligence platform in a classroom the artificial intelligence smart robotic platform moved around the room answered questions, in a clear voice and logged the faces of the students. I made the platform using a controller that is based on the ESP32. This ESP32-based controller is connected to a different things. It is connected to a motor driver, which is used for motion control. The ESP32-based controller is also connected to an ESP32-CAM module. This is used for capturing video in real time. Additionally the ESP32-based controller is connected to a MAX98357A amplifier and a speaker and these are used for giving feedback from the ESP32-based controller.

KEYWORDS: AI Robotics, Smart Campus, Facial Recognition, Attendance Automation, ESP32, ESP32-CAM, Large Language Model (LLM), Computer Vision, IoT-based Robot

I. INTRODUCTION

The rapid advancement in natural language processing (NLP) has led to the development of large language models (LLMs), which started with models such as BERT [2], GPT [3], [4], and Codex [5], and more recently, LLaMa [6], [7], Mistral [8] that are revolutionizing a wide range of applications. These models have achieved remarkable results in various tasks such as text generation, machine translation, and code synthesis, among others. A recent addition to this collection of models was the OpenAI ChatGPT [1], a pretrained generative text model which was finetuned using human feedback. In recent years there have been different attempts to incorporate language into robotics systems. These efforts have largely focused on using language token embedding models, LLM features, and multi-modal model features for specific form factors or scenarios. Applications range from visual-language navigation [9], [10], [11], language-based human-robot interaction [12], [13], and visual-language manipulation control [14], [15], [16]. However, despite the potential advantages of using LLMs in robotics, most of the existing approaches are restricted by a rigid scope and limited ChatGPT's ability to engage in a free-form dialog and capture long context allows users to interact with the model in a more natural fashion, with flexible behavior correction. At the same time, ChatGPT's extensive knowledge of mathematical constructs, geometry, commonsense afford it a much higher ability to understand and reason about the physical world compared to older LLM-based approaches which mainly focused on task planning that was dependent on a fixed set of functions and behaviors.

Motivation of the Project

The motivation behind developing **Robex** arises from the growing need for intelligent robotic systems capable of performing complex tasks with greater efficiency, accuracy, and autonomy. Traditional robotic systems are often limited by predefined programming and lack the ability to adapt to dynamic real-world environments. With rapid advancements in Artificial Intelligence, Machine Learning, Computer Vision, and IoT technologies, there is a strong opportunity to create a smart robotic platform that can think, learn, analyze, and respond intelligently.

Problem Statement

In today's rapidly evolving technological landscape, conventional robotic systems are limited in their ability to operate intelligently in dynamic and unpredictable environments. Most existing robots rely heavily on fixed programming, manual control, and predefined instructions, which restrict their adaptability, decision-making capability, and efficiency in real-world applications. These limitations reduce their effectiveness in critical sectors such as healthcare, industrial automation, surveillance, disaster response, and smart assistance systems.

II. OBJECTIVES

The main objective of the paper is to design and implement an intelligent robotic system that can operate autonomously, interact naturally with humans, and manage attendance automatically by leveraging artificial intelligence and computer vision technologies.

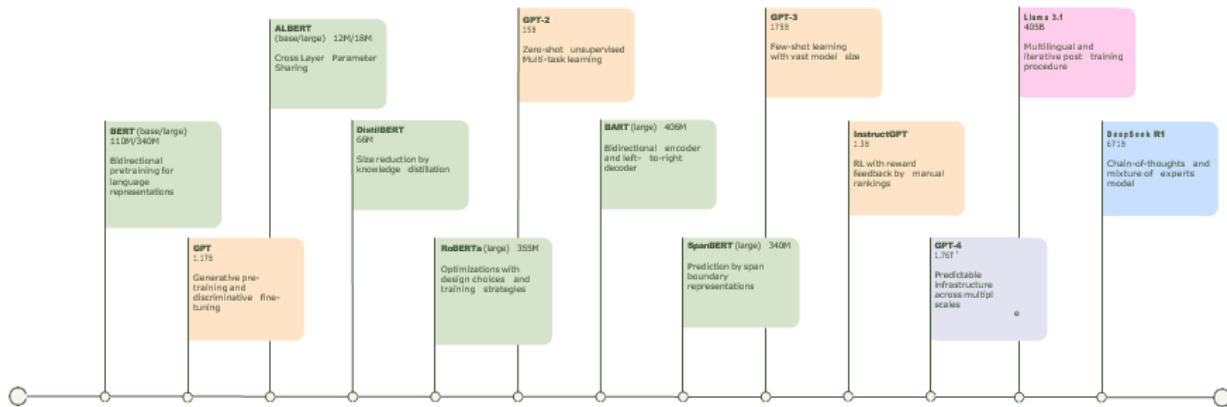
- To design and develop an AI-driven mobile robotic platform capable of controlled navigation using both manual inputs and voice-based commands in real time.
- To implement a natural language interaction mechanism using a large language model that enables the robot to interpret spoken instructions and generate appropriate spoken responses through speech synthesis.
- To develop a real-time visual perception module using an onboard camera for continuous video acquisition and face detection in dynamic and changing environments.
- To implement a facial recognition-based identification mechanism that enables accurate recognition of individuals without the need for physical contact or manual data entry.
- To design an automated attendance recording system that accurately logs entry and exit timestamps of recognized individuals in a cloud-hosted database.
- To integrate robot control, voice interaction, computer vision, and attendance management modules into a unified server-based architecture to ensure smooth coordination and data flow.
- To deploy the system on a cloud infrastructure that supports continuous 24×7 operation, remote access, and future scalability.
- To reduce the requirement for human involvement in routine monitoring and attendance-related tasks, thereby enhancing efficiency, accuracy, and reliability within educational environments.

III. METHODOLOGY

The proposed system is developed using a modular and layered methodology that integrates embedded hardware, computer vision, cloud computing, and web technologies. The overall workflow is divided into distinct functional stages to ensure scalability, reliability, and real-time performance.

In the proposed system, the robotic platform is developed using an ESP32-based microcontroller integrated with motor driver circuits, a camera module, and audio output components. The robot is configured to connect wirelessly to a centralized server through a local network or cloud infrastructure. This server functions as the primary processing unit responsible for handling control commands, executing artificial intelligence algorithms, and managing system data.

Robot navigation and task execution are controlled through user inputs provided via a web-based interface or spoken commands. Voice commands are captured at the user end and transmitted to the server, where they are processed using a generative artificial intelligence model for natural language understanding. Based on the interpreted intent, the server generates structured control instructions, which are then communicated back to the robot to enable real-time motion and task execution.



1. System Architecture Design

The complete system structure consists of three main components. Hardware Layer – Includes the mobile robotic platform, motor driver module, embedded controller, camera module, microphone, and audio output system. Processing Layer – The system contains a main server which performs artificial intelligence tasks and face recognition and attendance tracking and command processing. Cloud and Data Layer.

2. Hardware Implementation

The system uses a microcontroller-based robotic platform to achieve its physical movement capabilities. The embedded controller manages: Motor control for directional movement (forward, backward, left, right, stop). The system uses Wi-Fi to establish server communication. The system enables audio playback through a digital amplifier module. The system provides real-time camera streaming for vision-based processing. The robot establishes continuous server contact through HTTP requests which deliver motion commands and status information.

3. Voice Command Processing The system applies voice-based interaction through its client-server implementation. The user provides a voice command through the web interface. The server processes the natural language input using a large language model. The system transforms the interpreted command into structured JSON format which includes: Movement instruction Speech response The movement command is transmitted to the robot. The system uses a text-to-speech engine to convert the textual response into speech. The system streams the generated audio file to the robot which will play it back. The pipeline provides both real-time interaction capabilities and the ability to manage commands throughout the system.

4. Face Detection and Recognition Module

The attendance system operates through a computer vision pipeline which includes these three main stages: a) Face Detection The Haar Cascade classifier which has been pre-trained detects human faces from grayscale video frames that the embedded camera captures.

b) Face Dataset Collection For each new user registration: Multiple facial images are captured. The images undergo both resizing and normalization processes. Data is stored in structured directories labeled by user identity.

c) Model Training The Local Binary Pattern Histogram (LBPH) algorithm is employed for face recognition due to its robustness under varying lighting conditions. The system uses face images which users provide during registration to build the training dataset.

d) Face Recognition The system processes operations by The trained LBPH recognizer receives detected faces for analysis. Recognition validity depends on the confidence score. The system considers the individual recognized when the score falls below the defined threshold.

IV. REVIEW OF LITERATURE

Dahal et al. (2025)

Dahal et al. proposed AutoTA, a dynamic intent-based virtual teaching assistant for students using open-source LLMs like Vicuna and LLaMA. The study focuses on improving student learning support by enabling the assistant to grasp the intent behind student queries rather than just providing keyword-based answers. The system employs NLP techniques such as intent detection, context tracking, and response generation to deliver interactive academic support. AutoTA can manage repeated student questions, clarify concepts, and address contextual follow-up inquiries, making it

ideal for online learning. The authors also pointed out that open-source LLMs reduce reliance on paid commercial APIs and can be customized for specific educational areas. However, the study notes limitations, including reliance on dataset quality and fine-tuning methods.

Gogineni et al. (2025)

Gogineni et al. discussed techniques to fine-tune Large Language Models efficiently on limited hardware, particularly in single-GPU environments. The research explains that fine-tuning large models poses challenges due to memory needs, GPU power demands, and long training times. The authors suggested methods like memory optimization, power-aware GPU use, and training efficiency improvements. The paper highlights that these strategies make LLM fine-tuning possible for researchers and developers without access to expensive multi-GPU clusters. The study helps reduce costs in model training and supports scalable AI development on budget systems. Despite these benefits, the authors mention that scaling beyond single-GPU systems remains tough, and performance may decrease with very large models. Power-capping improves energy efficiency, but it can lengthen training times. Furthermore, fine-tuning results are highly dependent on GPU architecture, memory bandwidth, and system setup, which can make performance inconsistent across different hardware.

Abu-Jassar et al. (2023)

Abu-Jassar et al. developed a computer vision system for a small mobile humanoid robot in a smart environment. The system combines an ESP32-CAM module with Tiny-YOLO, a lightweight deep learning model suited for embedded vision tasks. The paper emphasizes real-time object recognition, object identification, and navigation support in crowded settings. The study shows that low-cost embedded hardware can enable robotic perception, allowing robots to detect objects, obstacles, and people without expensive camera systems. The study also points out that such vision systems are beneficial for smart homes, surveillance robots, and indoor service robots. However, due to the limited processing power of the ESP32-CAM, the robot cannot handle complex deep learning tasks on its own. The system requires external processing from a PC or server for heavier computations, increasing reliance on network connectivity. The model's accuracy drops in extreme lighting conditions like low or very bright light. Additionally, the performance may suffer in fast-moving situations due to frame rate restrictions.

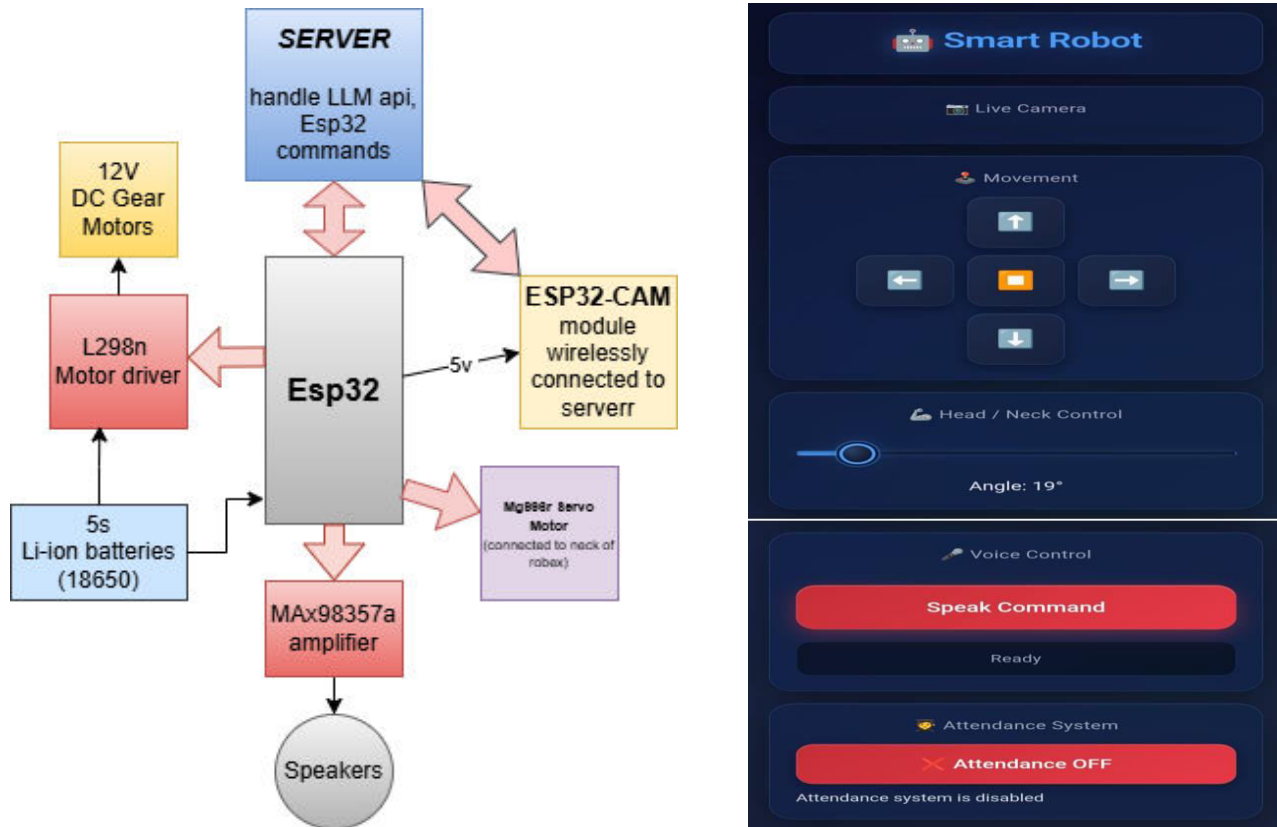
Zhou et al. (2022)

Zhou et al. introduced a 3D-printed soft robotic prosthetic hand (ACES-V2) equipped with embedded soft sensors for position and touch detection. The study aims to create a human-like robotic hand that can interact safely and flexibly. Soft sensors enhance tactile feedback and movement monitoring, improving object handling and grip control. The paper also points out that 3D printing allows for cost-effective and customizable prosthetic hand development, making it suitable for rehabilitation and assistive technology. The use of soft materials offers safer interaction than rigid robotic hands, minimizing the risk of injury and enhancing comfort.

MATERIALS USED

S.no.	Material	Description / Properties
1	ESP32	Central hardware controller for robot movement, Wi-Fi communication, and audio playback.
2	ESP32-CAM Module	Real-time image acquisition for face detection and recognition.
3	L298N Motor Driver	Control of DC motors for directional movement.
4	MAX98357A Digital Audio Amplifier + Speaker	Playback of synthesized voice responses.
5	Rechargeable Lithium-Ion Battery Pack	supply for robotic hardware.
5	Server-Side Framework	Python-based Web Framework (Flask).

OVERALL SYSTEM FRAMEWORK



V. CONCLUSION

The research work developed an intelligent robotic system which unites autonomous mobility voice-based interaction real-time computer vision and automated cloud-based attendance management into one comprehensive system. The system demonstrates how artificial intelligence together with embedded systems and Internet-enabled technologies creates an intelligent automated solution which works throughout academic settings. The proposed robotic platform enables real-time human-machine interaction through natural language processing and speech synthesis which allows users to execute commands with their voice without needing any technical skills. The system gains enhanced functionality through its face recognition module which automatically records attendance by tracking students' entry and exit times.

VI. FUTURE SCOPE

The new robotic system for taking attendance and talking to people is a starting point, for more research and making new technology. There are ways we can make the system better such as making it work more reliably and using it in more places. The future of this work includes:

- Integration of Deep Learning-Based Face Recognition
- Edge AI Processing
- Multi-Face and Crowd Analytics
- Biometric Multi-Factor Authentication
- Advanced Navigation and Autonomous Mapping
- Emotion and Behavior Analysis
- Cloud Dashboard with Analytics Visualization
- Mobile Application Integration
- Energy Optimization and Power Management

- Security and Data Encryption Enhancements
- Deployment in Other Domains
- Scalability for Large Campus Deployment
- AI-Based Voice Personalization
- Real-Time Notification System

REFERENCES

- [1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, pp. 511–518.
- [2] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [4] R. Szeliski, *Computer Vision: Algorithms and Applications*. London, UK: Springer, 2010.
- [5] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2010.
- [6] K. Ashton, "That 'Internet of Things' thing," *RFID Journal*, 2009.
- [7] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, 2016.
- [8] H. K. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.
- [9] D. Zhang, M. He, and J. Sun, "A cloud-based attendance management system using face recognition," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 5, pp



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details