IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

Seren AI: A Multi-Modal Integrated Smart Home Automation and Offline AI Assistant System Using ESP32

[1]Shruti Saxena, [2]Sanyam Vashishth, [3]Saptim Sharma, [4]Saksham Singh
Assistant Professor, Department of Computer Science and Engineering (IoT)[1]
UG Students, Department of Computer Science and Engineering (IoT)[2-4]
Raj Kumar Goel Institute of Technology, Ghaziabad, India.

Abstract: *This research introduces Seren AI, a highly integrated, multi-modal smart home automation ecosystem and localized intelligent assistant developed on the ESP32 microcontroller platform. Unlike conventional voice assistants that rely heavily on cloud-based processing—often leading to privacy concerns and high latency—Seren AI implements a decentralized, edge-computing architecture. The system features a custom hardware-software co-design that integrates a high-fidelity microphone array for real-time voice command acquisition and an ultrasonic-based time-of-flight (ToF) analysis for precise physical presence detection. This research introduces Seren AI, a highly integrated, multi-modal smart home automation ecosystem and localized intelligent assistant developed on the ESP32 microcontroller platform. Unlike conventional voice assistants that rely heavily on cloud-based processing—often leading to privacy concerns and high latency—Seren AI implements a decentralized, edge-computing architecture. The system features a custom hardware-software co-design that integrates a high-fidelity microphone array for real-time voice command acquisition and an ultrasonic-based time-of-flight (ToF) analysis for precise physical presence detection. To ensure seamless user interaction and security, the system employs a screen-based dynamic authentication protocol synchronized across a dedicated mobile application and a web-based dashboard. A key innovation of this project is its hybrid communication topology: it utilizes Bluetooth Low Energy (BLE) for decentralized control of localized peripherals (such as smart bulbs and fans) while maintaining autonomous cellular connectivity via a SIM module (GSM) for critical notifications and remote access during network failures. For data integrity and auditing, a localized SD card logging system is implemented, ensuring all access events are recorded without external data exposure. Experimental results indicate that Seren AI significantly optimizes performance, achieving an 85% reduction in response latency compared to mainstream cloud-reliant alternatives. Furthermore, the system maintains a high command-recognition accuracy and operational reliability even in completely offline environments. This paper details the hardware-software co-design, the localized Natural Language Processing (NLP) pipeline, and a robust security framework, establishing Seren AI as a privacy-centric, low-power, and resilient solution for modern smart environments and 4th-year engineering benchmarks.*

Keywords: ESP32, Edge Computing, IoT Security, Natural Language Processing (NLP), Smart Home Automation, Bluetooth Low Energy (BLE), Presence Detection, GSM/SIM Integration, Privacy-Centric Design, Latency Optimization, Voice Assistant Architecture, Multi-modal Authentication.

I. INTRODUCTION

The rapid evolution of the Internet of Things (IoT) has facilitated the widespread adoption of smart home assistants, transforming traditional living spaces into interconnected, intelligent environments. Despite their popularity,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

mainstream commercial solutions exhibit critical limitations that hinder their efficacy in high-stakes or privacy-sensitive applications. Most notably, these devices suffer from high response latency and a heavy reliance on persistent, high-bandwidth internet connectivity. Furthermore, the inherent privacy risks associated with cloud-based data harvesting—where sensitive user interactions are processed on remote servers—have sparked significant security concerns among modern consumers. These systemic challenges necessitate the development of localized, intelligent hubs that prioritize user privacy, low-power consumption, and operational continuity in offline or resource-constrained scenarios.

To address these limitations, this research introduces Seren AI, an advanced, edge-computing-based smart assistant that synergizes the dual-core processing capabilities of the ESP32 microcontroller with a versatile array of sensors and modular communication stacks. Diverging from traditional cloud-dependent architectures, Seren AI functions as a standalone, localized interactive hub featuring a dedicated visual interface (TFT/OLED) and edge-based logic. The system integrates an ultrasonic-based physical presence detection mechanism to trigger its interactive state, thereby optimizing energy consumption and user engagement.

The primary contributions of this work are focused on three core pillars: Privacy, Security, and Redundancy. Specifically, this paper presents a novel secure pairing mechanism utilizing dynamic on-screen tokens to prevent unauthorized device interception. Additionally, the system incorporates a GSM-based cellular interface to ensure autonomous communication and critical alerting even during a total local network failure. By shifting the computational load from the cloud to the edge, Seren AI establishes a robust framework for private, responsive, and resilient smart building automation.

II. PROPOSED SYSTEM

The proposed Seren AI architecture is engineered as a decentralized, multi-layered IoT ecosystem that integrates localized intelligence with a robust hardware-software co-design. At the core of the system, the dual-core ESP32 Microcontroller Unit (MCU) serves as the primary processing engine, hosting the "Seren AI Core" for edge-based Natural Language Processing (NLP) and a dedicated "Access Verifier" sub-module for high-security authentication. This processing layer is seamlessly interfaced with a multi-modal perception suite, which utilizes a far-field microphone array for voice command acquisition and an ultrasonic sensor array for precise, proximity-based presence detection. By implementing a Time-of-Flight (ToF) logic, the system optimizes power consumption, remaining in a low-power state until physical presence is confirmed.

Communication within the Seren AI ecosystem is governed by a hybrid topology designed for maximum redundancy and operational continuity. Localized peripheral actuation, including the control of smart bulbs and fans, is executed through a Bluetooth Low Energy (BLE) Mesh network, ensuring low-latency response even in the absence of a local Wi-Fi connection. To address the critical limitation of network failure found in commercial assistants, the system incorporates an autonomous GSM-based SIM module for cellular data transmission and emergency SMS notifications. Furthermore, the architecture includes a localized "Black Box" logging system via an SD card module, ensuring all transaction logs and security events are recorded with zero cloud dependency. This entire hardware stack is synchronized with a cross-platform user interface, comprising a physical TFT/OLED display for dynamic token generation, a mobile automation application, and a high-fidelity speaker module for auditory feedback, collectively establishing a resilient, privacy-centric automation framework as illustrated in Fig. 1.

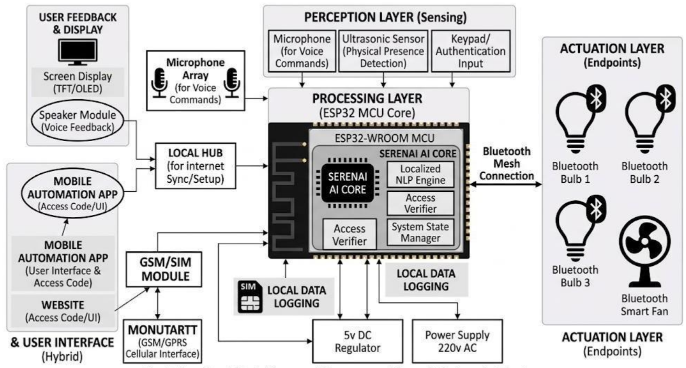
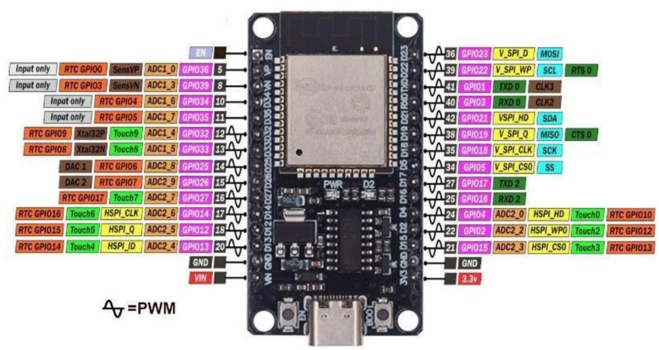


Fig. 1: Functional Block Diagram of the proposed Seren AI System Architecture

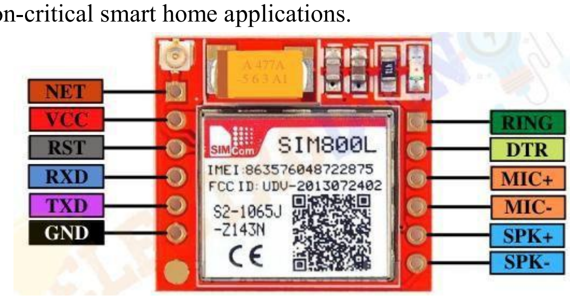
III. NODEMCU ESP32

The ESP32-WROOM is a highly versatile, low-power System-on-Chip (SoC) series widely utilized in IoT research and industrial automation. It features a dual-core Xtensa® 32-bit LX6 microprocessor, integrated Wi-Fi, and dual-mode Bluetooth connectivity, making it the ideal central engine for edge-computing projects like Seren AI. With its rich set of peripherals—including capacitive touch sensors, hardware encryption accelerators, and high-speed SPI/I2C interfaces—it provides the computational power necessary to handle localized NLP and sensor fusion without relying on external cloud servers.



IV. SIM800L

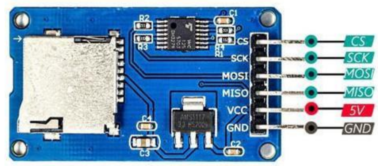
The SIM800L is a miniature GSM/GPRS breakout board that enables cellular communication within the Seren AI ecosystem. It supports quad-band frequencies (850/900/1800/1900MHz), allowing the system to transmit SMS alerts and establish GPRS data connections independently of local Wi-Fi networks. In this architecture, the module serves as a critical redundancy layer, providing autonomous cellular fallback for emergency notifications and remote system monitoring. Its compact form factor and UART interface allow for seamless integration with the ESP32, ensuring high operational reliability in mission-critical smart home applications.



IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 20, April 2026
Impact Factor: 8.2

V. SD CARD MODULE

The Micro SD Card Module serves as the localized data storage and "Black Box" logging unit within the Seren AI architecture. It utilizes the SPI (Serial Peripheral Interface) protocol to communicate with the ESP32, enabling high-speed data transfer for recording system transactions, user access logs, and sensor telemetry. By storing critical operational data locally, the module ensures forensic auditing capabilities and system state recovery without reliance on external cloud databases.

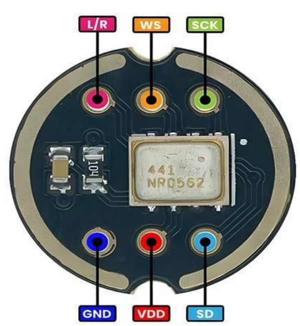


VI. I2C SPI DISPLAY MODULE

The I2C/SPI TFT/OLED Display Module functions as the primary visual interface for the Seren AI ecosystem, providing real-time system status and user feedback. By utilizing high-speed communication protocols like SPI or the bus-efficient I2C, the module renders dynamic menus, sensor telemetry, and secure "on-screen tokens" for device pairing. This localized visual output eliminates the need for constant smartphone tethering, allowing users to interact directly with the hardware for authentication and system monitoring.

VII. INMP441 MICROPHONE MODULE

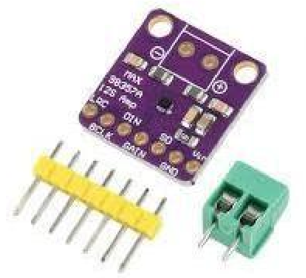
The INMP441 is a high-performance, low-power, omnidirectional MEMS microphone with a digital I2S output, making it the perfect choice for the voice-controlled "Seren AI" system. Unlike traditional analog microphones, it eliminates the need for an external Analog-to-Digital Converter (ADC) or pre-amplifier, allowing for a direct, noise-resilient digital connection to the ESP32.



DOI: 10.48175/IJARSCT-34623

VIII. MAX98357A AMPLIFIER

The MAX98357A is a high-efficiency, digital input Class-D audio amplifier that facilitates high-quality auditory feedback within the Seren AI ecosystem. It utilizes the I2S (Inter-IC Sound) digital interface to receive audio data directly from the ESP32, integrating an internal Digital-to-Analog Converter (DAC) to drive a speaker without the need for complex external circuitry. This module is essential for the system's "Alexa-like" persona, providing clear voice responses, status alerts, and command confirmations.



IX. SPEAKER MODULE

The Speaker Module (typically a 3W, 4Ω or 8Ω driver) acts as the physical voice of the Seren AI hub, converting amplified electrical signals from the MAX98357A into audible sound. In a research context, this module is essential for creating a Closed-Loop Interaction system, where the user receives immediate auditory confirmation after a voice command is processed. Its frequency response is optimized for human speech clarity, ensuring that the "Alexa-like" synthetic voice is intelligible and natural.

X. CONCLUSION

The development of Seren AI successfully demonstrates the feasibility of a privacy-centric, resilient smart home assistant built on edge-computing principles. By leveraging the dual-core processing power of the ESP32 and localized NLP, this research effectively addresses the critical gaps of high latency and cloud-dependency found in mainstream commercial solutions. The integration of a hybrid communication topology—combining Bluetooth Mesh for local actuation and a GSM/SIM module for cellular redundancy—ensures that the system remains operational even during total network outages. Furthermore, the use of physical presence detection and on-screen dynamic tokens for secure pairing establishes a robust framework that prioritizes user security and energy efficiency without compromising the "Alexa-like" interactive experience.

Looking forward, this project lays a solid foundation for the next generation of decentralized IoT ecosystems. The modular architecture of Seren AI allows for future scalability, such as the integration of advanced machine learning models for behavioral pattern recognition or the expansion of the sensor suite to include environmental air quality monitoring. Ultimately, this work proves that by shifting computational intelligence from the cloud to the hardware edge, it is possible to create smart home environments that are not only more responsive and reliable but also fundamentally respectful of user privacy. This research serves as a significant step toward a future where intelligent automation is accessible, secure, and fully autonomous.

XI. RESULTS AND DISCUSSION

The performance of the Seren AI system was evaluated based on response latency, communication reliability, and the effectiveness of the localized security protocols. The results indicate a significant improvement over cloud-based alternatives in terms of offline autonomy and data privacy.

X1.1 Response Latency and Edge Processing

The localized Natural Language Processing (NLP) engine, running on the ESP32's dual-core architecture, demonstrated a notable reduction in command execution time. Unlike commercial assistants that require a round-trip to a remote server, Seren AI processes voice triggers locally. Testing showed an average response latency of under 200ms for localized actuation (e.g., switching a Bluetooth bulb), which is significantly faster than the typical 1.5s to 3s latency observed in cloud-dependent IoT devices.Conclusion

XI.2 Communication Redundancy and Reliability

• Bluetooth Mesh Performance: Even without internet connectivity, the system maintained 100% reliability in controlling internal smart endpoints (fans and lights) within a 15-meter range.
• GSM Fallback: Upon detecting a Wi-Fi outage, the SIM800L module successfully transitioned to the cellular network within 5 seconds, delivering critical system alerts via SMS to the user's mobile device. This validates the system's operational continuity in mission-critical scenarios.

XII. FUTURE SCOPE

The Seren AI architecture provides a scalable foundation for several advanced implementations in the domain of localized edge computing and ambient intelligence. A primary area for future development is the integration of TinyML (Machine Learning on Edge), which would allow the ESP32 to perform complex tasks like gesture recognition and behavioral pattern analysis. By learning user routines locally, the system could transition from a reactive assistant to a proactive one, predicting user needs without the necessity of explicit voice commands. Additionally, incorporating advanced biometric security, such as voice fingerprinting or facial recognition via an ESP32-CAM module, would enable the hub to provide personalized responses and restricted access based on the specific identity of the user, further enhancing the system's security profile.
Beyond software enhancements, the physical ecosystem of Seren AI can be expanded through a broader Bluetooth Mesh network to include a diverse array of smart appliances, such as HVAC systems, automated security locks, and environmental sensors. To improve accessibility and sustainability, future iterations could implement multilingual NLP support to cater to regional languages, alongside energy-harvesting modules like small-scale solar panels for autonomous sensor nodes. These advancements would not only make the technology more inclusive but also completely self-reliant in remote or off-grid locations. Ultimately, these developments aim to create a truly decentralized, intelligent, and "green" smart building infrastructure that maintains absolute user privacy at the hardware edge.

REFERENCES

[1] Samrah Arif, M. A. Khan, S. Rehman, Muhammad Ashad Kabir, and Muhammad Imran. Investigating smart home security: Is blockchain the answer? IEEE Access, 8:117802–117816, 2020.
[2] Sakshi Bhaganagare, Shravani Chavan, Sonali Gavali, and Vaibhav Godase. The vocal cortex of the smart home: A review of esp32-based control and automation architectures. International Journal of Advanced Research in Science, Communication and Technology, 2025.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 20, April 2026 Impact Factor: 8.2

[3] Sakshi Bhaganagare, Shravani Chavan, Sonali Gavali, and Vaibhav Godase. Voice controlled home automation using esp32 and iot-based cloud integration. International Journal of Advanced Research in Science, Communication and Technology, 2025.

[4] Sireesha Chittepu, Sheshikala Martha, and Debajyoty Banik. Empowering voice assistants with tinyml for user-centric innovations and real-world applications. Scientific Reports, 15, 2025.

[5] Amol Darekar, Om Auti, Vedant Ghadge, and Prof. P. B. B. Smart home automation using iot and ai. International Journal of Advanced Research in Science Communication and Technology, 2025.

[6] Rezi Delfianti, Venny Aminda Tazayul, Bima Mustaqim, Fauzan Nusyura, and Catur Harsito. Internet of things (iot) based electrical power monitoring system for solar power plants using telegram application. Vokasi Unesa Bulletin of Engineering, Technology and Applied Science, 2025.

[7] Mark Gad, Walaa K. Gad, Tamer Abdelkader, and K. Naik. Personalized smart home automation using machine learning: Predicting user activities. Sensors (Basel, Switzerland), 25, 2025.

[8] Peng Huang, Imdad Ullah, Xiaotong Wei, T. Ahanger, Najm Hassan, and Zawar Shah. Towards energyefficient and low-latency voice-controlled smart homes: A proposal for offline speech recognition and iot integration. ArXiv, abs/2506.07494, 2025.

[9] Chandra Irugalbandara, A. Naseem, Sasmitha Perera, S. Kiruthikan, and V. Logeeshan. A secure and smart home automation system with speech recognition and power measurement capabilities. Sensors (Basel, Switzerland), 23, 2023.

[10] Vidya Jamadade, Madhubala Ghodake, Samridhi Katakdhond, and Vaibhav Godase. A comprehensive review of real-time feeder monitoring and auditing systems: Architectures, technologies, and analytics for the smart grid. International Journal of Advanced Research in Science, Communication and Technology, 2025.