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 5, Issue 8, April 2025
Impact Factor: 7.67

# An ESP32-Based Voice Assistant Integrating Gemini AI and Mobile Speech Recognition

**Prof.VikasDesai[1], Amol Waghmare[2], Sandesh Shinde[3], Pritam Rangari[4], Ganesh Shejul[5], Dishant Thakur[6], Suraj Valunj[7]**
Professor, Department of Information Technology[1]
Students, Department of Information Technology[2-7]
AISSMS Institute of Information Technology, Pune, India

**Abstract:** *This work proposes the development and implementation of a low-cost, fully customizable voice assistant that uses the ESP32 microcontroller, Google's Gemini AI for natural language processing, and a mobile app built using MIT App Inventor for speech recognition. The system supports users to interact with the voice assistant going through verbal commands, processed locally on a mobile device, and sent to the ESP32 for AI response generation and audio playback. Through the convergence of these disparate technologies, the resultant platform exhibits flexibility, scalability, and accessibility, rendering it appropriates for a broad list of voice-controlled applications within Internet of Things (IoT) systems.The architecture utilizes the ESP32 as the central controller to manage Wi-Fi communication, information exchange with the Gemini AI API, and control of audio playback via an I2S digital-to-analog converter module. The mobile app executes speech recognition utilizing Android's inbuilt speech services and sends the transcribed text to the ESP32 via HTTP POST requests. After being processed by the Gemini AI, the text response is synthesized to speech and played back to the user. This method delegates intensive computation to the mobile device and cloud AI services, making the solution feasible for resource-limited embedded systems [1].This work explores system design, hardware-software integration, implementation complexities, performance analysis, and possible improvements thereby adding to the body of literature on embedded AI and mobile-enabled smart systems. The results indicate that hybrid mobile-embedded voice assistants provide an optimal compromise among capability, cost-effectiveness, and expandability for emerging IoT applications. The proposed solution in this research is not only feasible from a technical perspective but also addresses the crucial issues like cost, development simplicity, and versatility for practical application, thus rendering it an appealing option for prototyping and academic purposes. Moreover, performance metrics gathered under stringent testing—like a 94% command recognition rate, an end-to-end latency of 1.8 seconds on average, and more than 12 hours of uninterrupted operation on a 2200 mAh Li-Ion battery—support the system's real-world usability and power efficiency.*

**Keywords:** ESP32, Voice Assistant, Gemini AI, MIT App Inventor, IoT, Mobile Speech Recognition, Embedded



DOI: 10.48175/IJARSCT-25577