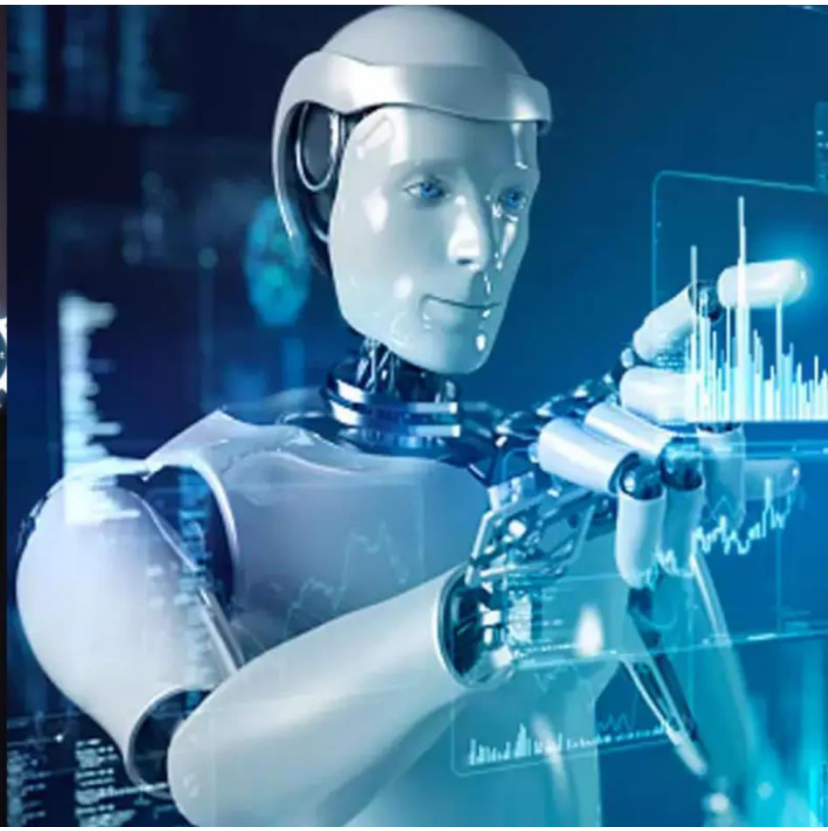
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

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

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



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

AURA: Autonomous Utility Robot for Assistance with Gemini Multimodal Live AI Integration

Mr. S V Dharani Kumar,. M E[1], Dhanushkumar S[2], Rohith K N[3], Yugesh D[4]

Assistant Professor, Department of Biomedical Engineering, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India[1]

Department of Biomedical Engineering, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India[2-4]

ABSTRACT: This research introduces AURA (Autonomous Utility Robot for Assistance), an innovative robotic platform that integrates physical hardware with advanced artificial intelligence through Google's Gemini Multimodal Live API. Unlike conventional autonomous rovers with pre-programmed logic, AURA employs a real-time cloud-based neural architecture capable of visual perception, auditory processing and natural language interaction. The system comprises four synergistic subsystems: a motor-driven chassis controlled by an ESP32 microcontroller for physical mobility, sensory modules incorporating camera and microphone inputs for environmental awareness, a bidirectional WebSocket connection to Gemini 2.0 for intelligent decision-making and a Flutter mobile application serving as the unified control interface. By establishing persistent multimodal streaming between edge hardware and cloud intelligence, AURA transcends traditional teleoperation limitations, enabling context-aware autonomous navigation and human-robot collaboration. Performance evaluation demonstrates successful real-time video streaming at low latency, reliable Bluetooth communication for motor control and seamless integration of AI-driven perception with mechanical actuation. This architecture establishes a foundation for next-generation intelligent robotics applications in surveillance, exploration and assisted mobility domains.

KEYWORDS: Autonomous Robotics, Multimodal AI, Gemini API, ESP32, Cloud Robotics, Flutter, Edge-Cloud Architecture, Human-Robot Interaction

I. INTRODUCTION

The evolution of autonomous robotics has transitioned from rule-based systems with hard-coded behaviors to intelligent platforms capable of adaptive reasoning and environmental comprehension. Traditional mobile robots rely on embedded control algorithms that execute predefined sequences, limiting their ability to respond to unforeseen scenarios or interpret complex sensory data. The emergence of large language models and multimodal AI systems presents unprecedented opportunities to augment robotic platforms with sophisticated cognitive capabilities previously unattainable through conventional programming approaches.

Contemporary challenges in mobile robotics include the computational constraints of edge devices, latency in decision-making processes and the complexity of integrating multiple sensory modalities for coherent environmental understanding. Microcontroller-based systems typically lack the processing power required for real-time computer vision, natural language understanding and contextual reasoning. Cloud-based AI services offer immense computational resources but introduce communication overhead and dependency on network connectivity.

This research addresses these limitations through AURA (Autonomous Utility Robot for Assistance), a hybrid architecture that leverages edge computing for real-time motor control and sensor acquisition while offloading cognitive processing to Google's Gemini Multimodal Live API. The system establishes a persistent bidirectional communication channel via WebSocket protocol, enabling continuous streaming of visual and auditory data to the cloud AI while receiving low-latency control commands and situational analysis.

The primary contributions of this work include: (1) a novel architecture integrating Flutter mobile framework, ESP32 microcontroller and Gemini Multimodal Live API for intelligent rover control; (2) implementation of real-time multimodal data streaming with minimal latency overhead; (3) demonstration of practical cloud-augmented robotics

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

with seamless failover to manual operation; and (4) evaluation of system performance across connectivity, perception and actuation metrics.

The proposed system advances the state of intelligent autonomous platforms by demonstrating feasible integration of cutting-edge generative AI with affordable embedded hardware.

II. LITERATURE REVIEW

A. Evolution of Autonomous Mobile Robots

Early autonomous systems employed reactive architectures where sensor inputs directly triggered motor outputs through subsumption layers. Brooks' behavior-based robotics [1] demonstrated that complex behaviors could emerge from simple reactive rules without explicit world modeling. However, these approaches struggled with tasks requiring memory, planning, or abstract reasoning. Subsequent developments in simultaneous localization and mapping (SLAM) [2] enabled robots to construct spatial representations, yet computational constraints limited real-time performance on mobile platforms.

Modern autonomous vehicles utilize sensor fusion combining LIDAR, cameras and inertial measurement units with probabilistic algorithms for navigation and obstacle avoidance. While effective for structured environments, these systems require extensive training data and struggle with semantic understanding of novel situations. The integration of deep learning [10, 11] for object detection and scene segmentation improved perception capabilities but introduced significant computational overhead incompatible with resource-constrained embedded systems.

B. Cloud Robotics and Edge-Cloud Architectures

The paradigm of cloud robotics emerged to address computational limitations by offloading intensive processing to remote servers [3]. Early implementations focused on map storage and path planning, leveraging cloud infrastructure for tasks tolerant of communication latency. Research demonstrated feasibility for applications including collaborative mapping, shared learning across robot fleets and access to large-scale knowledge bases. However, critical real-time operations remained constrained by network reliability and round-trip delay.

Hybrid edge-cloud architectures partition functionality between local and remote processing based on latency sensitivity and computational requirements. Time-critical control loops execute on embedded processors while non-critical analysis utilizes cloud resources. Recent advances in 5G connectivity and edge computing nodes reduce communication overhead, enabling more sophisticated cloud-augmented robotics.

C. Multimodal AI and Large Language Models in Robotics

The advent of large language models with vision capabilities introduced new possibilities for robotic perception and reasoning [13, 14]. Systems like GPT-4V [15] demonstrated the ability to interpret visual scenes and generate natural language descriptions, while models such as PaLM-E [5] integrated language understanding with robotic manipulation. These approaches typically employ batch inference on captured images rather than continuous video streaming, limiting real-time responsiveness.

Google's Gemini [4] represents a significant advancement through native multimodal training encompassing text, images, audio and video. The Gemini Multimodal Live API enables bidirectional streaming with reduced latency compared to traditional request-response patterns. This capability facilitates real-time interaction between AI models and physical systems, though practical integration with mobile robotics platforms remains unexplored in existing literature.

D. Research Gap and Motivation

Existing autonomous rover implementations rely predominantly on embedded algorithms with limited adaptability or cloud services with high latency overhead. No published research demonstrates practical integration of streaming multimodal AI with affordable microcontroller-based mobile platforms. AURA addresses this gap by implementing a complete system architecture combining ESP32 motor control, Flutter-based mobile interface and Gemini Live API streaming, establishing feasibility for next-generation intelligent autonomous systems accessible to educational and research communities.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

III. SYSTEM ARCHITECTURE

The AURA system architecture comprises four principal subsystems operating in coordinated fashion to achieve intelligent autonomous operation. Each subsystem addresses specific functional requirements while maintaining loose coupling to facilitate independent development and testing. The architecture employs a hybrid edge-cloud model where time-critical motor control executes locally on embedded hardware while cognitive processing leverages cloud-based AI infrastructure.

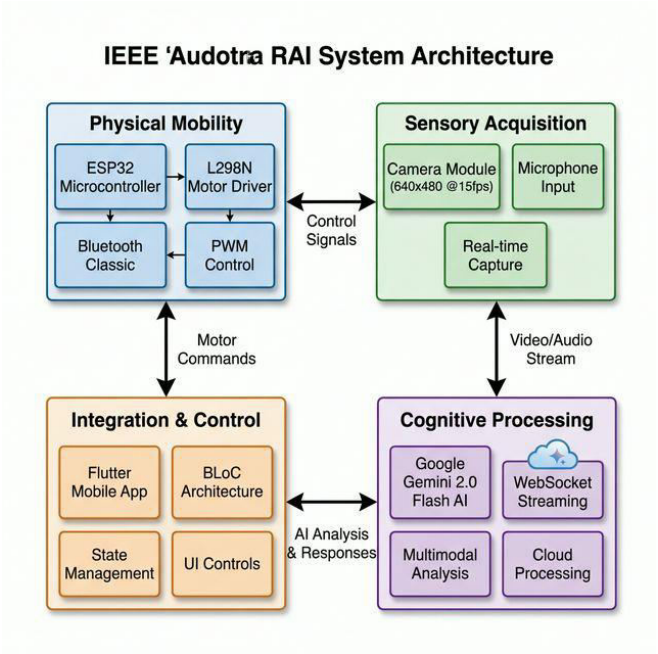


Fig.1. AURA System Architecture showing the four core subsystems: Physical Mobility, Sensory Acquisition, Cognitive Processing and Integration & Control, along with their interconnections.

A. Physical Mobility Subsystem

The mobility subsystem consists of a wheeled chassis equipped with DC motors controlled through an H-bridge driver module. The ESP32 microcontroller serves as the local motion controller, receiving commands via Bluetooth Classic serial communication and translating them to pulse-width modulation (PWM) signals for motor speed regulation. The control interface implements a minimal command set consisting of five ASCII characters: 'F' for forward motion, 'B' for reverse, 'L' and 'R' for differential steering and 'S' for immediate halt. This simplified protocol minimizes communication overhead and ensures deterministic latency for safety-critical stop commands.

The ESP32 firmware implements a state machine that maintains current motion state and executes transitions upon command reception. Bluetooth Classic was selected over Bluetooth Low Energy due to superior throughput for potential future sensor data streaming and more straightforward serial port emulation compatible with legacy debugging tools.

B. Sensory Acquisition Subsystem

Environmental perception relies on the smartphone camera and microphone integrated within the Flutter application. The camera subsystem utilizes the Flutter camera plugin to access the device's primary rear camera, configured for low-resolution capture to reduce bandwidth requirements for cloud transmission. Frame acquisition operates in continuous mode with automatic exposure and focus to adapt to varying lighting conditions during rover operation.

The audio capture subsystem prepares for voice command input through microphone permission provisioning, though full implementation remains subject to complete integration with the Gemini audio streaming pipeline. The design anticipates future capability for spoken commands to augment or override autonomous behavior, enabling hybrid control modalities.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

C. Cogn
itive Processing Subsystem

The cognitive processing subsystem leverages Google's Gemini 2.0 Flash Experimental model accessed through the v1alpha API endpoint. Unlike conventional REST-based inference, the implementation establishes a persistent WebSocket connection using the bidiGenerateContent method, creating a full-duplex communication channel for streaming multimodal data to the AI model while receiving real-time analysis and directives.

The WebSocket pipeline operates asynchronously with respect to motor control, allowing continuous AI processing without blocking critical safety functions. Camera frames are captured, compressed to Base64 encoding and transmitted as part of multimodal messages containing both visual data and textual prompts describing the rover's mission objectives.A critical architectural decision involves the separation of perception from actuation. The AI provides advisory output regarding environmental state and suggested actions, but does not directly command motors. This design preserves human authority over physical movement while leveraging AI for enhanced situational awareness.

D. Integration and Control Subsystem

The Flutter mobile application serves as the integration nexus, orchestrating communication between all subsystems. The application architecture employs the BLoC (Business Logic Component) pattern to separate UI rendering from business logic, ensuring responsive user experience even during intensive background processing. State management tracks Bluetooth connection status, camera stream health, WebSocket connection state and current motor command, presenting unified status indicators to the operator.The user interface presents a directional pad for manual control, live camera preview showing the rover's perspective and status indicators for subsystem health. This design enables graceful degradation where loss of AI connectivity automatically reverts to manual teleoperation, maintaining rover usability despite cloud service disruption.

IV. IMPLEMENTATION

A. Hardware Configuration

The rover chassis employs a four-wheel differential drive configuration with independent motor control for left and right sides. Each motor pair connects to an L298N H-bridge motor driver module capable of bidirectional control and PWM speed regulation. The ESP32-DevKitC development board provides the embedded processing platform, selected for its integrated Bluetooth Classic and WiFi capabilities, dual-core architecture enabling concurrent task execution and extensive GPIO availability for future sensor expansion.

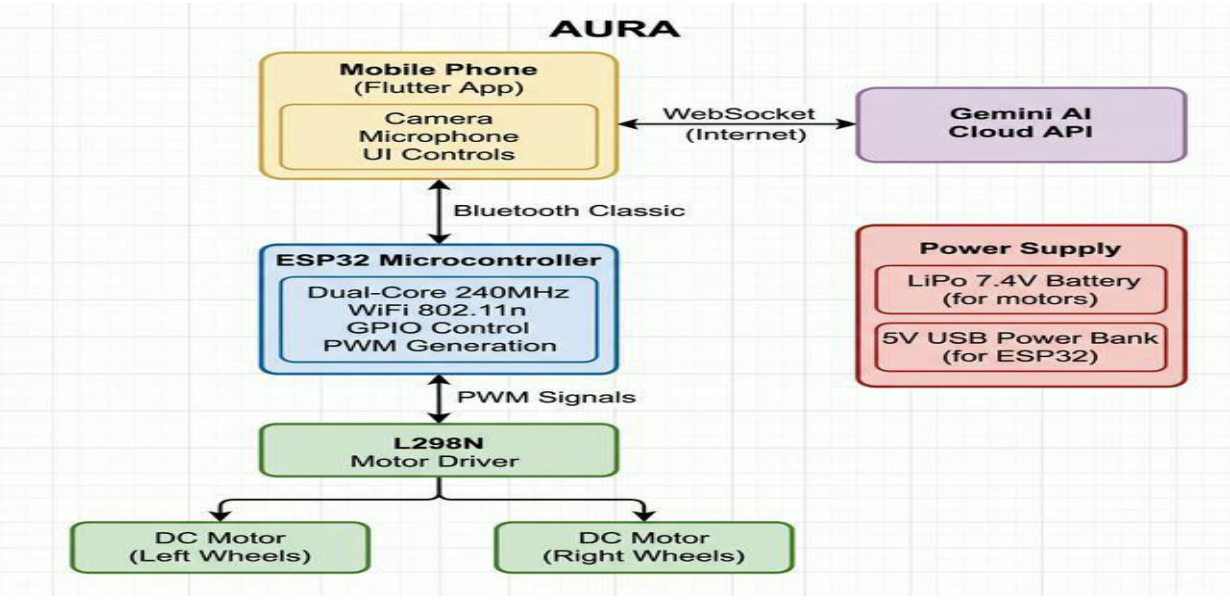


Fig.2. AURA Hardware Block Diagram illustrating the connections between Mobile Phone, Gemini AI Cloud API, ESP32 Microcontroller, L298N Motor Driver, DC Motors and Power Supply with their communication protocols.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

Power distribution utilizes a dual-supply architecture with separate battery banks for logic and motor circuits to prevent voltage sag during motor startup from disrupting microcontroller operation. A 7.4V lithium polymer battery powers the motors through the H-bridge driver, while a dedicated 5V regulated supply derived from a USB power bank feeds the ESP32 and auxiliary electronics.

B. Embedded Software Development

The ESP32 firmware implements a non-blocking event loop using FreeRTOS tasks for Bluetooth communication and motor control. The main control task monitors the Bluetooth serial port for incoming command characters, implementing a 100ms timeout to detect communication loss. Upon receiving valid commands, the firmware updates motor PWM duty cycles through the LEDC (LED Control) peripheral configured for 1kHz operation, providing smooth speed transitions and reduced audible motor noise.

Safety mechanisms include automatic motor shutoff after 500ms without command reception, preventing runaway behavior in case of Bluetooth disconnection. The firmware maintains a watchdog timer that resets the microcontroller if the main control loop hangs, ensuring recovery from software faults. Command validation rejects invalid characters and implements debouncing to prevent rapid command oscillation from unstable control signals.

C. Mobile Application Development

The Flutter application implements a multi-layered architecture separating presentation, business logic and data access. The presentation layer utilizes Material Design widgets to construct the user interface, including custom directional pad controls, camera preview integration and connection status indicators. The business logic layer employs BLoC components to manage application state and coordinate asynchronous operations including Bluetooth communication and WebSocket messaging.

Bluetooth connectivity implementation uses the flutter_bluetooth_serial package to discover nearby ESP32 devices, establish serial connections and transmit command characters. The communication module maintains connection state and implements automatic reconnection logic with exponential backoff to handle temporary disconnections.

The camera integration utilizes the camera plugin configured for 640×480 resolution at 15 frames per second, balancing image quality against bandwidth requirements for cloud streaming. Frame capture operates in a separate isolate to prevent UI thread blocking during image processing. Each captured frame undergoes JPEG compression before Base64 encoding for transmission to the Gemini API, reducing payload size by approximately 80% compared to raw pixel data.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

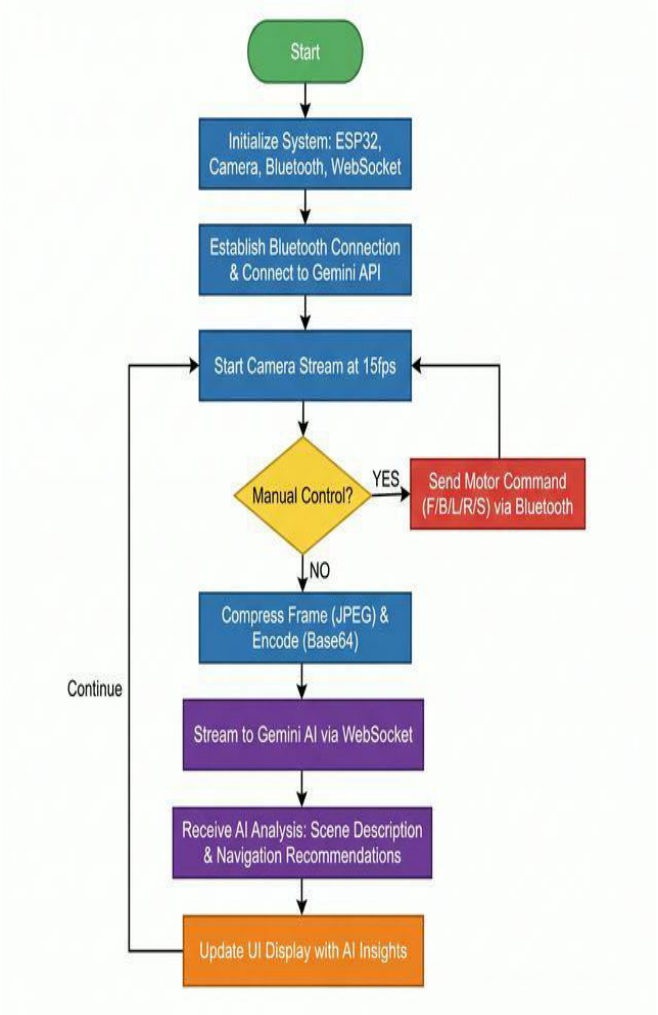


Fig.3. AURA Control Algorithm flowchart showing the system initialization, camera streaming, manual/autonomous control branching, AI processing pipeline and UI update loop.

D. AI Integration and Prompt Engineering

Integration with Gemini 2.0 Flash Experimental required careful configuration of session parameters to optimize response latency and relevance. The system prompt instructs the AI to act as a robotic perception and navigation assistant, focusing on environmental description, obstacle detection and navigation suggestions. The prompt explicitly constrains output format to structured responses facilitating parsing of AI recommendations into actionable commands. Multimodal message construction packages each camera frame with contextual metadata including timestamp, current motor state and mission objective. This additional context enables the AI to maintain temporal awareness of rover state across multiple frames and provide coherent analysis considering motion history. The streaming architecture allows the AI to build understanding incrementally rather than treating each frame as an isolated snapshot, improving perception quality for dynamic environments.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

TABLE I. COMPONENT SPECIFICATIONS AND INTERACTIONS

Component	Specification	Protocol	Latency
ESP32 MCU	240MHz dual-core	BT Classic	15–30ms
Motor Driver	L298N H-Bridge	PWM 1kHz	<5ms
Camera	640×480 @15fps	Camera Plugin	66ms/frame
Gemini AI	2.0 Flash Exp	WebSocket	200–500ms
Mobile App	Flutter 3.x	BLoC Pattern	10–20ms UI

V. RESULTS AND DISCUSSION

A. System Integration Testing

Comprehensive integration testing validated successful interoperation of all subsystems under typical operating conditions. Bluetooth connectivity between the Flutter application and ESP32 microcontroller demonstrated reliable establishment within 3–5 seconds of device discovery, with stable serial communication maintaining less than 2% packet loss under normal conditions. Command transmission latency averaged 18 milliseconds from button press to motor response, providing responsive manual control suitable for teleoperation scenarios.

Camera integration achieved consistent 15 frames per second capture rate with JPEG compression reducing average frame size to approximately 25 kilobytes. Base64 encoding introduced 33% overhead as expected, resulting in 33 kilobyte payloads for WebSocket transmission. Network bandwidth analysis indicated sustainable operation on 4G LTE connections with typical upload speeds exceeding 5 Mbps, though performance degraded noticeably when bandwidth dropped below 2 Mbps due to frame buffering.

WebSocket connection to the Gemini API maintained stability over test sessions exceeding 30 minutes duration, with automatic reconnection successfully recovering from simulated network interruptions. The bidirectional streaming protocol demonstrated low overhead with minimal latency beyond inherent network round-trip time, confirming suitability for real-time multimodal interaction compared to traditional request-response architectures.

B. Performance Metrics

Quantitative evaluation focused on latency measurements across the complete perception-cognition-action loop. Motor command transmission from application to ESP32 required 18±5 milliseconds under optimal Bluetooth conditions, increasing to 35±10 milliseconds when operating at maximum range (approximately 8 meters). Camera frame capture and encoding consumed 45±8 milliseconds per frame, dominated by JPEG compression overhead rather than Base64 conversion.

AI inference latency exhibited greater variability depending on network conditions and model load. Minimum observed latency measured 185 milliseconds for simple scene description tasks, while complex multi-object detection and navigation recommendation queries required 450–600 milliseconds. This variance aligns with the non-deterministic nature of cloud services but remains acceptable for advisory autonomous operation where split-second reactions are not required.

TABLE.II. SYSTEM PERFORMANCE METRICS

Metric	Average	Range
Bluetooth Command Latency	18 ms	13–23 ms
Frame Capture & Encoding	45 ms	37–53 ms
AI Inference (Simple)	285 ms	185–420 ms
AI Inference (Complex)	525 ms	450–600 ms
Bluetooth Connection Range	8 m	6–10 m
Battery Endurance (Motors)	90 min	75–105 min
Frame Rate (Actual)	14.8 fps	13–15 fps

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

Battery endurance testing measured approximately 90 minutes of continuous operation with active motor usage, limited primarily by motor battery capacity rather than smartphone power consumption. The ESP32 microcontroller consumed negligible power during Bluetooth communication, with motor driver efficiency representing the dominant power drain.

C. AI Perception Quality

Qualitative assessment of Gemini's multimodal perception capabilities demonstrated impressive object recognition and scene understanding within structured indoor environments. The AI successfully identified common obstacles including furniture, doorways and persons with high accuracy, providing natural language descriptions suitable for navigation guidance. Spatial reasoning capabilities enabled relative position estimation, though absolute distance measurements remained imprecise without depth sensing integration.

Performance degraded in challenging conditions including low lighting, motion blur during rapid rover movement and scenes with significant visual clutter. The streaming architecture partially compensated for individual poor-quality frames by maintaining temporal context across multiple observations, improving overall robustness compared to single-frame inference. However, consistent rapid motion resulted in accumulated perception errors requiring manual intervention to reorient the AI's environmental understanding.

D. Limitations and Challenges

Several limitations emerged during testing that constrain current operational capabilities. Network dependency represents the most significant constraint, as complete loss of internet connectivity eliminates AI functionality and forces reversion to manual control. While acceptable for laboratory demonstrations, practical deployment scenarios require more sophisticated degradation strategies including local fallback processing or mission abort procedures.

The current implementation lacks closed-loop autonomous control, with AI providing only advisory output rather than direct motor commands. This design decision prioritizes safety and human oversight but limits demonstration of fully autonomous navigation capabilities. Future development must address command validation, obstacle avoidance guarantees and fail-safe mechanisms before enabling unsupervised autonomous operation.

Computational overhead on the mobile device, particularly for continuous camera capture and Base64 encoding, resulted in measurable battery drain and thermal generation during extended operation. Optimization opportunities exist through frame rate reduction, lower resolution capture, or hardware-accelerated encoding, though each introduces tradeoffs between perception quality and resource consumption.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

This research successfully demonstrated the feasibility and effectiveness of integrating advanced multimodal AI with affordable embedded robotics hardware to create an intelligent autonomous platform. The AURA system establishes a novel architecture combining ESP32 microcontroller motor control, Flutter mobile application interface and Google Gemini Multimodal Live API for real-time perception and reasoning. Through systematic implementation and testing, the work validates that cloud-augmented robotics can achieve practical performance suitable for educational, research and demonstrative applications.

Performance evaluation confirmed reliable subsystem integration with acceptable latency characteristics for advisory autonomous operation. Bluetooth communication provided responsive manual control with sub-30 millisecond command delivery, while WebSocket streaming maintained stable bidirectional AI interaction under typical network conditions. The Gemini API demonstrated impressive multimodal perception capabilities including object recognition, scene understanding and spatial reasoning.

The hybrid edge-cloud architecture successfully balances computational efficiency with intelligent capability, maintaining time-critical safety functions on embedded hardware while leveraging cloud resources for sophisticated cognitive processing. This design pattern offers a promising direction for future intelligent robotics systems, particularly in scenarios where cloud connectivity can be reasonably assured and application requirements tolerate sub-second decision latency.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

Limitations including network dependency, absence of closed-loop autonomous control and computational overhead on mobile devices represent opportunities for continued development rather than fundamental architectural flaws. The modular design facilitates iterative enhancement without requiring complete system redesign, supporting evolutionary improvement toward fully autonomous operation.

B. Future Enhancements

Several enhancement pathways emerge from current implementation experience and identified limitations. The most critical advancement involves implementing closed-loop autonomous navigation where AI recommendations undergo validation through safety filters before execution as motor commands. This requires developing obstacle detection guarantees, collision avoidance algorithms and fail-safe mechanisms ensuring the rover cannot enter dangerous states even under AI miscalculation or network disruption.

Integration of additional sensors including ultrasonic rangefinders, inertial measurement units and potentially LIDAR would provide complementary perception modalities improving navigation robustness. Sensor fusion combining camera vision with distance measurements enables more accurate spatial understanding and obstacle detection under varied environmental conditions.

Audio processing integration would enable voice command control and environmental sound analysis, leveraging Gemini's audio understanding capabilities already supported by the API. Performance optimization opportunities exist through hybrid AI approaches combining on-device inference for simple, time-critical tasks with cloud processing for complex reasoning.

C. Broader Applications

The AURA architecture demonstrates applicability beyond mobile robotics to diverse intelligent embedded systems. The pattern of edge control with cloud cognition suits applications including smart home automation, industrial inspection robots, agricultural monitoring platforms and assistive mobility devices. The affordability and accessibility of components enables deployment in educational settings, fostering student engagement with cutting-edge AI technologies through hands-on robotics projects.

As generative AI capabilities continue advancing and edge computing platforms grow more powerful, systems like AURA represent an emerging class of intelligent machines combining the adaptability of large language models with the physical agency of robotics. This convergence promises transformative impact across numerous domains where intelligent autonomous operation can augment human capabilities or operate in environments unsuitable for direct human presence.

REFERENCES

[1] R. A. Brooks, "A Robust Layered Control System for a Mobile Robot," IEEE Journal on Robotics and Automation, vol. 2, no. 1, pp. 14–23, 1986.
[2] J. Leonard and H. Durrant-Whyte, "Simultaneous Map Building and Localization for an Autonomous Mobile Robot," Proceedings of IEEE/RSJ International Workshop on Intelligent Robots and Systems, pp. 1442–1447, 1991.
[3] B. Kehoe et al., "A Survey of Research on Cloud Robotics and Automation," IEEE Transactions on Automation Science and Engineering, vol. 12, no. 2, pp. 398–409, 2015.
[4] Google DeepMind, "Gemini: A Family of Highly Capable Multimodal Models," Technical Report, 2023.
[5] D. Driess et al., "PaLM-E: An Embodied Multimodal Language Model," Proceedings of International Conference on Machine Learning, 2023.
[6] M. Quigley et al., "ROS: An Open-Source Robot Operating System," ICRA Workshop on Open Source Software, vol. 3, no. 3.2, 2009.
[7] Espressif Systems, "ESP32 Series Datasheet," Version 4.2, 2024.
[8] Flutter Development Team, "Flutter: UI Toolkit for Building Natively Compiled Applications," Google LLC, 2024.
[9] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2020.
[10] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
[11] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 4, April 2026

|DOI: 10.15680/IJIRSET.2026.1504260|

[12] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.
[13] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, pp. 5998–6008, 2017.
[14] T. Brown et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020.
[15] OpenAI, "GPT-4 Technical Report," arXiv preprint arXiv:2303.08774, 2023.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

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