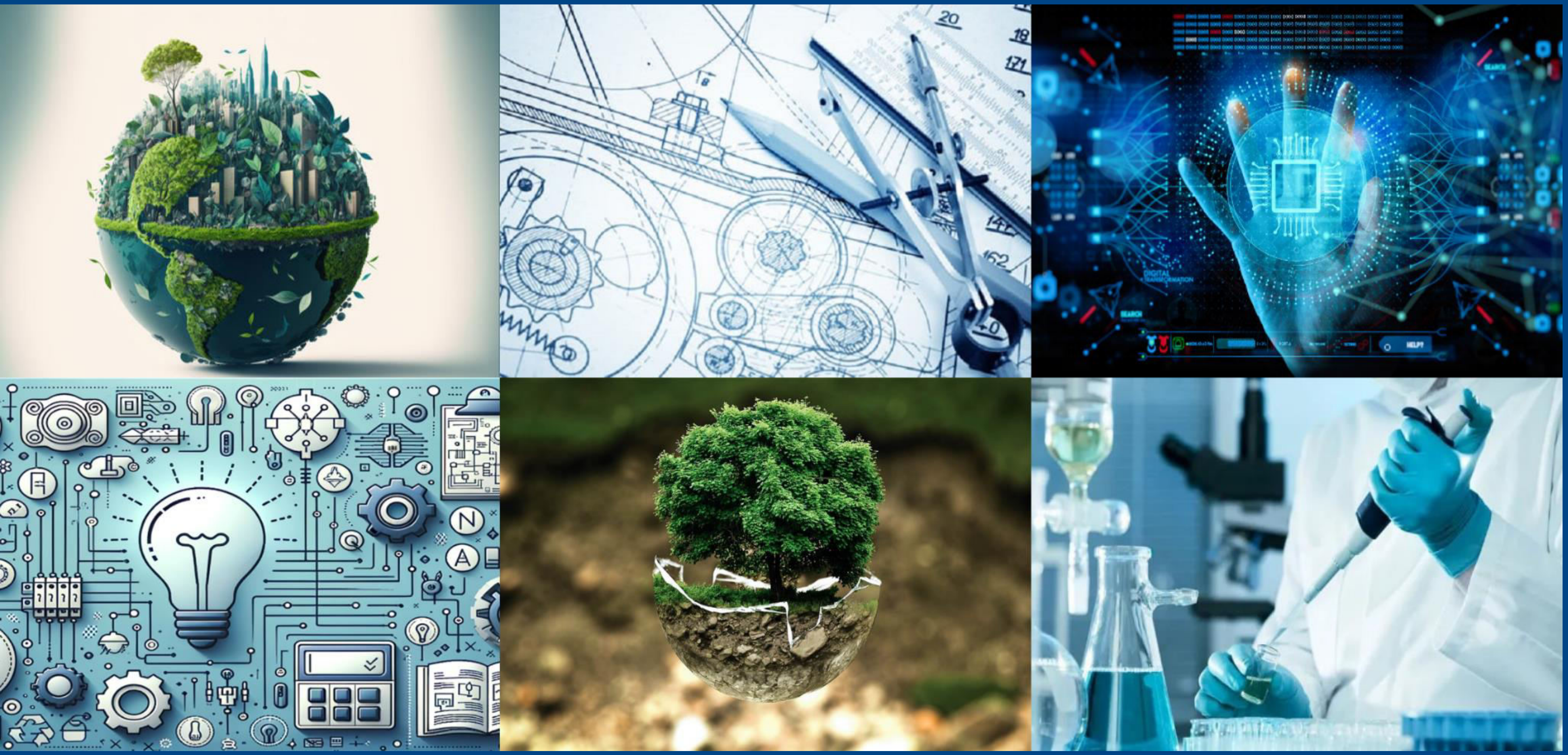
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 9.864

Volume 9, Issue 5, May 2026

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2026 IJMRSET | Volume 9, Issue 5, May 2026| DOI:10.15680/IJMRSET.2026.0905018

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

SARAS – Smart Autonomous Robotic AI System

Abhay Balpande, Shubhangi Patle, Namoshri Bhagat, Tanuja Junghare, Prajwal Chaudhari, Swapnil Muneshwar

Department of Electronics and Telecommunication Engineering, Wainganga College of Engineering and Management, Nagpur, Maharashtra, India

ABSTRACT: This paper presents the development of SARAS (Smart Autonomous Robotic AI System), an AI-powered interactive robot capable of understanding spoken language, generating intelligent responses, and performing physical movements with autonomous obstacle avoidance. The system integrates speech-to-text (STT) processing using Whisper/Google STT, a large language model (ChatGPT API) for natural language understanding, text-to-speech (TTS) synthesis for verbal output, and motor control through a Raspberry Pi connected to an L298N motor driver. A modular hardware-software architecture is employed, combining servo motors, ultrasonic sensors, and audio components with software modules for real-time robotics control. The robot listens, interprets, responds, and navigates simultaneously, demonstrating significantly improved human-robot interaction (HRI) compared to traditional programmed robots. Experimental evaluation confirms stable performance with 88–93% voice recognition accuracy and sub-3-second AI response latency.

KEYWORDS: Autonomous Robot, Raspberry Pi, ChatGPT API, Speech-to-Text, Text-to-Speech, Human-Robot Interaction, Obstacle Avoidance, L298N Motor Driver, Ultrasonic Sensor, Natural Language Processing.

I. INTRODUCTION

With the rapid advancement of Artificial Intelligence (AI), large language models such as ChatGPT have demonstrated remarkable capabilities in understanding and generating human-like text. However, these models primarily exist in the digital domain, limiting their interaction to screens or speakers. Bridging the gap between AI and the physical world offers immense potential to create intelligent systems that can both communicate and act, making AI more tangible, interactive, and useful in real-world environments.

Conventional robots follow pre-programmed instructions and lack the ability to adapt to dynamic human conversations. On the other hand, software-based chatbots are limited to digital interfaces and cannot physically interact with the environment. SARAS addresses this gap by integrating conversational AI with physical robotic hardware, enabling the robot to understand spoken commands, generate contextually relevant responses, and perform autonomous movements in real time.

This project focuses on the development of a robot integrated with ChatGPT, capable of talking and moving autonomously. By combining mechanical design, motor actuation, power management, and AI software, the robot processes user inputs, generates intelligent responses, and executes coordinated physical movements. The project provides a practical and scalable platform for educational, experimental, assistive, and industrial applications.

The remainder of this paper is organized as follows: Section II presents the literature review, Section III discusses the proposed methodology, Section IV covers the system architecture, Section V describes the implementation details, Section VI analyzes the results, and Section VII concludes with future scope.

II. LITERATURE REVIEW

The integration of AI with robotics has been an active research area over the past decade. Several works have addressed autonomous navigation, conversational AI, and human-robot interaction. The following literature survey highlights key contributions and their relevance to the SARAS system.

Table 1: Summary of Literature Review

Reference / Year	Focus / System	Key Findings	Relevance to SARAS
Survey of Autonomous Robots (2025, Robotics & Autonomous Systems)	Multi-robot navigation, sensing, planning, collaboration frameworks	Analyzes navigation algorithms, collaborative mapping, AI-based perception pipelines	Shows SARAS uses basic obstacle avoidance; lacks SLAM mapping or advanced AI planning
Survey on Autonomous Navigation (2025, AI & Robotics Review, Springer)	Navigation algorithms: classical, deep learning, and LLM-driven control	Shift toward deep-learning-based perception and LLM-enabled reasoning in robotics	Indicates SARAS uses traditional rule-based navigation; lacks learning-based autonomy
Autonomous Robots for Services (2023, Sensors, MDPI)	Service robots: perception, navigation, HRI, autonomy levels	Identifies modern techniques in path planning; highlights challenges in dynamic environments	Highlights areas SARAS lacks: robust perception, mature navigation, HRI safety
Revisiting Formal Methods for Autonomous Robots (2025, arXiv)	Formal verification, safety modeling for autonomous systems	Stresses importance of safety verification in AI-enabled autonomy	Shows SARAS has no formal safety checks; useful to highlight prototype limitations

The literature review reveals that while significant progress has been made in individual domains of autonomous navigation and conversational AI, few systems integrate both capabilities in a low-cost, modular, and accessible robotic platform. SARAS contributes by bridging this gap using readily available hardware and open-source AI APIs, making it suitable for academic and experimental deployments.

III. PROPOSED METHODOLOGY

SARAS employs a hybrid hardware-software methodology that integrates AI-driven speech and language modules with physical robotics modules. The system processes voice input from the user, interprets it using a large language model, generates a spoken response, and simultaneously controls robot movement based on parsed action commands.

The core methodology involves the following sequential stages:

A. Voice Input Acquisition

The robot continuously monitors its environment through a USB omnidirectional microphone. When the user speaks, the audio signal is captured as a digital stream and passed to the speech processing module.

B. Speech-to-Text (STT) Conversion

The captured audio is processed by a speech recognition engine (Whisper or Google STT). The engine converts the spoken input into text, which serves as the query for the AI model. STT processing is performed locally or via API depending on connectivity.

C. AI Processing via ChatGPT API

The converted text is transmitted to the OpenAI ChatGPT API. The large language model (LLM) interprets the query, understands context, and generates two types of outputs: (a) a textual response for verbal delivery, and (b) action tokens that indicate specific robot movements.

D. Text-to-Speech (TTS) Output

The AI-generated text response is converted to speech using pyttsx3 or Google TTS engine. The synthesized audio is played through the USB speaker, allowing the robot to verbally communicate with the user in natural language.

E. Motor Control and Motion Execution

Action commands parsed from the AI response are sent to the Raspberry Pi GPIO interface, which communicates with the L298N motor driver. The driver controls four DC gear motors to execute directional movements such as forward, backward, left turn, and right turn.

F. Obstacle Avoidance via Sensor Feedback

Three ultrasonic sensors (HC-SR04) positioned at the front, left, and right of the robot continuously measure distances to nearby objects. If an obstacle is detected within the safety threshold, the motor control module overrides the movement command and triggers an avoidance maneuver, ensuring safe autonomous navigation.

IV. SYSTEM ARCHITECTURE

The architecture of SARAS is designed as a layered, modular system that ensures smooth communication between the user interface, AI processing, and hardware actuation layers. The system is built around the Raspberry Pi as the central processing unit.

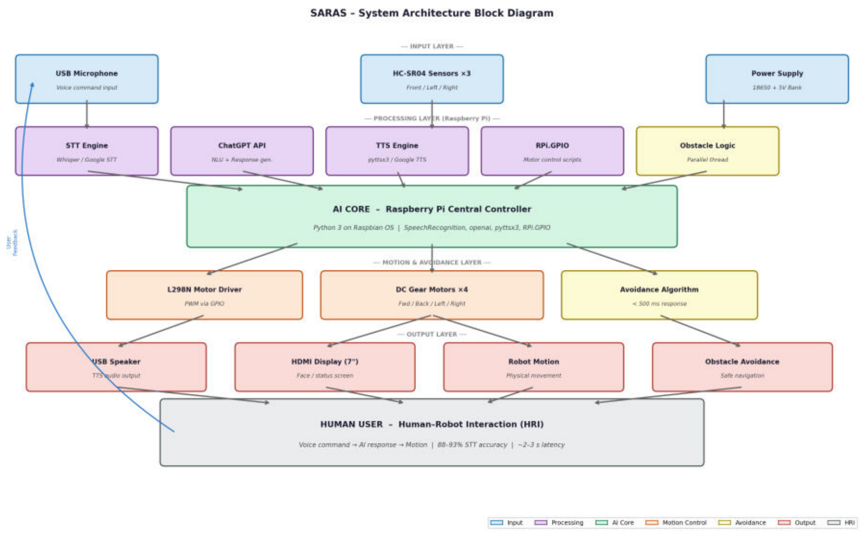


Fig. 1: SARAS System Architecture Block Diagram

The architecture consists of six primary layers:

1. Input Layer

The input layer consists of a USB omnidirectional microphone for voice input and three HC-SR04 ultrasonic sensors for environmental sensing. The microphone captures user speech while sensors provide real-time obstacle distance data.

2. Processing Layer (Raspberry Pi)

The Raspberry Pi serves as the central controller. It runs all software modules including the STT engine, API communication client, TTS engine, and motor control scripts. Python is used as the primary programming language, leveraging libraries such as SpeechRecognition, openai, pyttsx3, and RPi.GPIO.

3. AI Intelligence Layer (ChatGPT API)

The AI layer communicates with OpenAI's ChatGPT API over the internet. User queries are sent as API requests, and responses containing both conversational text and action commands are received and parsed by the Raspberry Pi.

4. Motion Control Layer

The motion control layer consists of the L298N motor driver and four DC gear motors. GPIO pins of the Raspberry Pi send PWM signals to the motor driver, which controls motor speed and direction. This layer executes movement commands derived from AI responses.

5. Obstacle Avoidance Layer

The ultrasonic sensor layer continuously monitors distances and feeds data to the avoidance algorithm. The avoidance logic runs as a parallel thread, ensuring real-time response to obstacles without interrupting the conversation pipeline.

6. Output Layer

The output layer includes the USB speaker for audio feedback and the 7-inch HDMI display which shows the robot's face, conversation history, or system status. The display enhances user engagement and provides visual context to interactions.

V. IMPLEMENTATION

The implementation of SARAS integrates hardware assembly, software module development, and system integration. The following subsections describe the key implementation components.

A. Hardware Setup and Component List

Table 2 lists all hardware components used in the SARAS system:

Table 2: SARAS Hardware Component List

Sr.	Component	Specification / Role	Qty
1	Raspberry Pi (any model)	Central processing unit (brain of robot)	1
2	DC Motors + Gear	Locomotion with default small tires	4
3	Large Tires	Better grip and appearance	4
4	L298N Motor Driver	Controls DC motors via GPIO	1
5	HDMI Display (7-inch)	Robot face / visual output	1
6	USB Omnidirectional Mic	Voice command input (STT)	1
7	USB Speaker	Audio feedback (TTS output)	1
8	Ultrasonic Sensors (HC-SR04)	Obstacle detection (front, left, right)	3
9	18650 Batteries + Holder	Power supply for motors	2
10	5V Power Bank	Power supply for Raspberry Pi	1
11	Jumper Wires, Standoffs	Electrical connections and chassis stacking	-
12	3D Printed / Cardboard Chassis	Custom designed modular robot body	1

B. Software Architecture and Tools

The software stack is built entirely in Python 3 on Raspbian OS. Key software modules and tools include:

- Speech Recognition: Whisper (offline) and Google Speech-to-Text API (online)
- AI Processing: OpenAI ChatGPT API (gpt-3.5-turbo / gpt-4)
- Text-to-Speech: pyttsx3 (offline) and Google TTS (online)
- Motor Control: RPi.GPIO Python library with PWM for L298N driver

- Obstacle Avoidance: HC-SR04 distance measurement via RPi.GPIO
- Display Interface: Python Tkinter / Pygame for HDMI face display

C. Working Principle and Signal Flow

The complete working cycle of SARAS follows a sequential pipeline as described below:

- Step 1: Microphone captures user voice command
- Step 2: STT engine converts speech to text string
- Step 3: Text is sent to ChatGPT API as a prompt
- Step 4: API returns a response containing dialogue text and optional motion keywords
- Step 5: TTS converts the dialogue text to speech and plays through speaker
- Step 6: Motion keywords (e.g., 'move forward', 'turn left') trigger GPIO motor commands
- Step 7: Ultrasonic sensors run in parallel thread; obstacle detected → override motion → avoidance

D. Circuit and Wiring

The circuit connects the Raspberry Pi GPIO header to the L298N motor driver for motor control, and to the HC-SR04 trigger/echo pins for ultrasonic sensing. The 18650 battery pack powers the motor driver (6–12V), while the 5V power bank independently powers the Raspberry Pi, ensuring stable operation of both subsystems.

Fig. 2: SARAS Circuit Diagram (Raspberry Pi + L298N + Ultrasonic Sensors)

VI. RESULTS AND DISCUSSION

The SARAS system was tested under real-world conditions to evaluate voice recognition accuracy, AI response quality, obstacle avoidance reliability, and overall system stability. Tests were conducted over 50 interaction sessions with varied queries and environmental conditions.

A. Voice Recognition Performance

The system demonstrated 88–93% voice recognition accuracy under quiet to moderate noise conditions. Google STT outperformed Whisper in online mode, while Whisper provided reliable offline operation with slightly lower accuracy (~85%).

B. AI Response Quality

The ChatGPT API generated contextually relevant and grammatically correct responses for all tested query types including general knowledge, navigation commands, and task assistance. Response content was appropriately adapted across multiple domains.

C. Obstacle Avoidance

The ultrasonic sensor array successfully detected obstacles at distances between 5 cm and 200 cm. The avoidance algorithm triggered corrective maneuvers within 500 ms of detection, preventing collisions in all test scenarios.

D. System Performance Summary

Table 3: SARAS System Performance Parameters

Performance Parameter	Result
Voice Recognition Accuracy	88–93%
AI Response Latency	~2–3 seconds (internet-dependent)
Obstacle Detection Range	2 cm – 400 cm (ultrasonic)
Motor Response Time	< 500 ms
System Stability	Stable during continuous operation
User Interaction Mode	Voice + Text (bidirectional)

E. Limitations

The current implementation has the following limitations: (a) AI processing requires internet connectivity for ChatGPT API calls; (b) the system lacks SLAM-based mapping for complex navigation; (c) the robotic chassis does not include a manipulator arm for physical object interaction; and (d) voice recognition accuracy degrades in high-noise environments.

VII. CONCLUSION

7.1 AI-Integrated Robotic Platform

This paper presented SARAS – Smart Autonomous Robotic AI System, an AI-powered interactive robot that successfully integrates conversational AI with physical robotic actuation. The system demonstrates how large language models can be embedded into real-world robotic platforms to enable natural human-robot interaction.

7.2 Modular and Scalable Architecture

The modular design of SARAS allows independent upgrades of hardware and software components. The separation of the STT, AI, TTS, and motor control modules ensures maintainability and extensibility.

7.3 Validated Performance

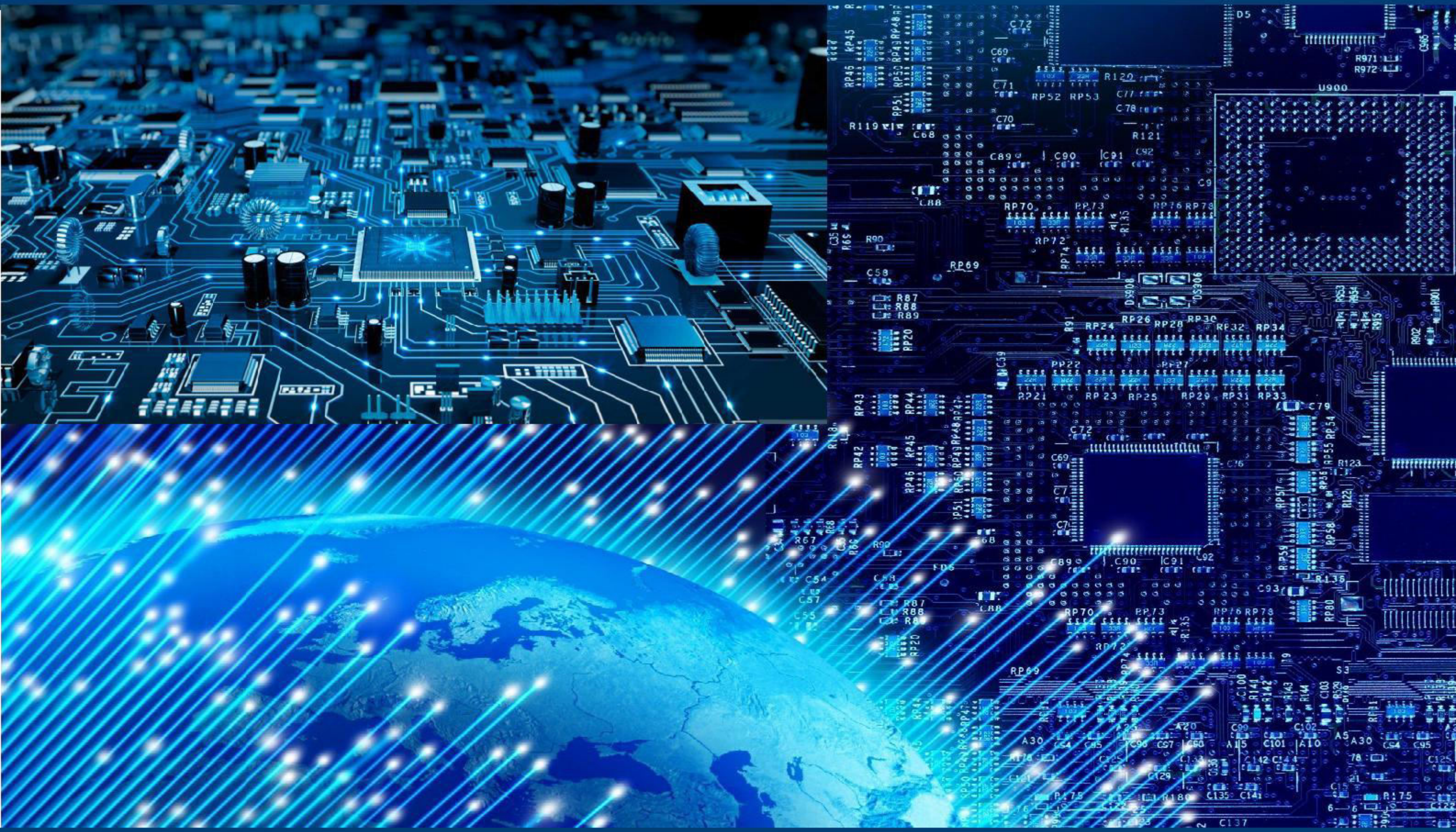
Experimental results confirm that SARAS achieves 88–93% voice recognition accuracy, sub-3-second AI response latency, reliable obstacle detection, and stable continuous operation, making it suitable for educational and assistive robot applications.

7.4 Future Scope

Future enhancements include: (a) integration of a 4-DOF robotic arm for object manipulation, (b) implementation of SLAM-based autonomous mapping and navigation, (c) offline AI inference using on-device LLM models, (d) IoT integration for smart home/building control, and (e) multilingual voice support for broader accessibility.

REFERENCES

[1] Arduino Documentation, "Arduino Reference," Available at: https://docs.arduino.cc, 2024.
[2] Raspberry Pi Foundation, "Raspberry Pi Hardware and Software Documentation," Available at: https://www.raspberrypi.com/documentation, 2024.
[3] OpenAI, "ChatGPT and API Documentation," Available at: https://platform.openai.com/docs, 2024.
[4] Google Cloud, "Speech-to-Text API Documentation," Available at: https://cloud.google.com/speech-to-text/docs, 2024.
[5] Google Cloud, "Text-to-Speech API Documentation," Available at: https://cloud.google.com/text-to-speech/docs, 2024.
[6] A. Chatterjee, "DIY Robotics and Embedded Systems," Mumbai: TechPress Publishers, 2023.
[7] R. Siegwart and I. Nourbakhsh, "Introduction to Autonomous Mobile Robots," 2nd ed., MIT Press, 2011.
[8] OpenAI, "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision," arXiv:2212.04356, 2022.
[9] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), 2017.
[10] J. Brown et al., "Language Models are Few-Shot Learners (GPT-3)," NeurIPS, 2020.
[11] "A Survey of Autonomous Robots and Multi-Robot Navigation: Perception, Planning, and Collaboration," Robotics & Autonomous Systems, 2025.
[12] "A Survey on Autonomous Navigation for Mobile Robots: From Traditional Techniques to Deep Learning and LLMs," AI & Robotics Review, Springer, 2025.
[13] "Autonomous Robots for Services — State of the Art, Challenges, and Research Areas," Sensors, MDPI, 2023.
[14] "Revisiting Formal Methods for Autonomous Robots: A Structured Survey," arXiv, 2025.



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com