



AI-Driven Personal Assistant Robot Integrated with IoT

A Low-Cost, Modular Bipedal Platform for Natural Interaction and Smart Home Integration

¹Dr. Sneha Tirth, ²Samyak Dhole, ³Abhinav Shukla, ⁴Yash Gangan, ⁵Ritesh Rathod

¹Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹Department of Computer Engineering

¹Trinity College of Engineering and Research, Pune, Maharashtra, India

Abstract: This paper presents the design and implementation of a mobile, AI-driven personal assistant robot that integrates intelligent movement, natural interaction, and seamless connectivity using Wi-Fi. The robot is built on a lightweight 3D-printed PLA chassis and powered by servo motors—DS3230 for the lower body and MG995 for the upper body—offering a practical and cost-effective mobility platform. Locomotion is managed by Central Pattern Generators (CPGs), enabling smooth and stable bipedal walking, even with payloads close to the robot's own weight (~6–7 kg). The robot understands voice commands using the Cohere API for natural language processing and detects human presence and gestures using the YOLO object detection model. User interaction is made intuitive through real-time speech and visual input. IoT functionality is handled over Wi-Fi, allowing the robot to interact with smart home devices and services. In lab tests, the robot achieved reliable path following with an average positional error of ~3 cm, responsive gesture recognition within 30 ms, and accurate voice understanding in varied conditions. This work demonstrates the feasibility of building accessible, human-centric assistant robots using modern AI tools and common hardware. Future development will focus on enhancing real-world adaptability, social interaction, and long-term usability in home environments.

Index Terms - Personal Assistant Robot, Bipedal Locomotion, Human–Robot Interaction, Internet of Things (IoT), Modular Robot Design, Smart Home Automation, Artificial Intelligence.

I. INTRODUCTION

With the rising interest in service robotics, there has been a growing push to develop intelligent personal assistant robots that can help people in daily life. Unlike industrial robots which work in structured and predictable environments, assistant robots are designed to operate in homes—spaces that are often unstructured, dynamic, and personal. These robots are expected not just to perform tasks but to understand and adapt to human needs, offering both utility and companionship.

In recent years, improvements in artificial intelligence and IoT technologies have made it possible to design robots that can understand speech, recognize visual inputs, and interact with smart home devices. Deep learning-based perception models like YOLO have significantly improved real-time object detection, while natural language processing (NLP) tools such as the Cohere API make it possible for robots to understand and respond to voice commands in a conversational manner. Combined with Wi-Fi connectivity, these features allow robots to communicate with IoT-enabled devices like lights, thermostats, and appliances, enabling a more integrated smart home experience.

This paper presents the design and development of an AI-powered personal assistant robot that blends practical mechanical design with advanced software intelligence. The robot is built using a lightweight, 3D-printed PLA chassis and utilizes servo motors—DS3230 in the lower body and MG995 in the upper body—to enable flexible bipedal movement. The system is powered by a 5200 mAh, 11.1V LiPo battery, making it portable and suitable for indoor environments. At a total mass of around 6–7 kg, the robot strikes a balance between structural robustness and mobility.

For locomotion, we adopt Central Pattern Generators (CPGs), a biologically inspired approach that generates rhythmic joint movements similar to how humans and animals walk. This technique allows the robot to walk stably on two legs without needing complex control logic or external balancing aids.

In terms of user interaction, the robot supports natural communication via voice. Commands are processed through an automatic speech recognition engine and interpreted using the Cohere API, allowing the robot to handle tasks such as "turn on the living room lights", etc. On the vision side, YOLO is used to detect people and objects, helping the robot recognize users and navigate its environment more effectively.

IoT integration is handled through Wi-Fi, allowing the robot to interact with smart devices seamlessly. Whether turning off lights or responding to triggers from other connected devices, the robot acts as both a personal helper and a smart home hub.

In essence, this work demonstrates how a human-friendly, AI-driven assistant robot can be built using accessible components and open tools. Our design emphasizes real-world usability, ease of interaction, and meaningful integration with home environments—all while keeping the technical framework straightforward and replicable.

II. HARDWARE ARCHITECTURE

2.1 Actuation and Structure

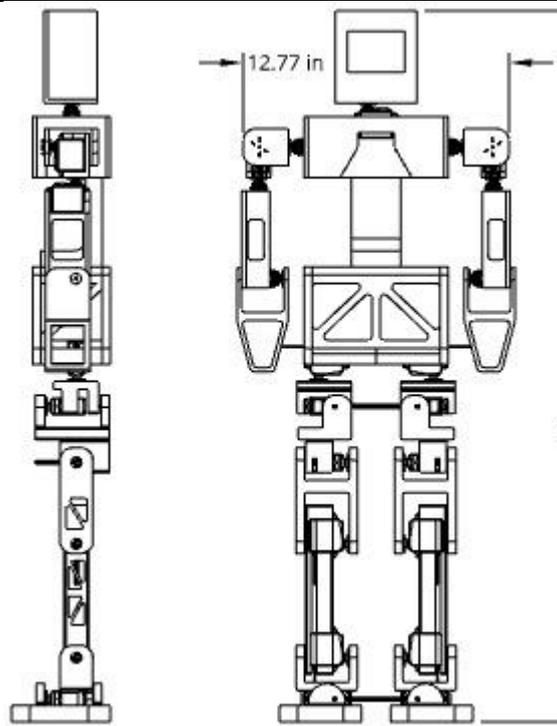
The robot's drive system is based on high-torque servo actuators, specifically the DS3230 and MG995, which are installed at each joint to control limb movement. These servo motors offer a reliable and cost-effective solution for precise joint positioning, making them suitable for bipedal locomotion in lightweight personal robots. Instead of complex brushless motors or hydraulic systems used in research-grade humanoids, we opted for digital servo actuators that balance simplicity, efficiency, and ease of integration.

All structural components of the robot's body are 3D-printed using PLA, a lightweight yet sufficiently rigid material ideal for prototyping. This approach allows for easy customization and quick iteration while keeping the overall mass of the robot under 7 kg. The internal frame, brackets, and covers are all designed to minimize weight without compromising mechanical stability.

To sense orientation and balance, the robot uses the BNO055 IMU, which provides fused 9-axis data (accelerometer, gyroscope, magnetometer).

This allows the control system to adjust the robot's posture and maintain stability during walking or standing still.

The servo motors are controlled using a PCA9685 driver board, which provides PWM signals to each actuator. While torque feedback is not available, motion is monitored through predefined control loops and the robot's current orientation, allowing for stable gait generation through Central Pattern Generators (CPGs). The end result is a compact, efficient robot platform capable of dynamic bipedal motion and responsive interaction within home environments.



(Figure 1.0 - Robot's Physical Structure)

The physical structure of the robot is based on a modular, full-body bipedal frame fabricated entirely using 3D-printed PLA components. The design emphasizes both functional segmentation and mechanical simplicity, allowing for rapid prototyping, ease of assembly, and future scalability. As shown in Figure 1.0, the robot features a broad suite of components well-suited for indoor navigation and interaction in domestic environments.

The frame is divided into distinct subassemblies including the head unit, torso, upper limbs, and lower limbs. Each segment is independently attachable via standardized mounting interfaces, facilitating part replacement, maintenance, and experimentation with different sensor and actuator configurations. The symmetrical leg design incorporates knee and ankle joints driven by DS3230 servo motors, with reinforced cavities to house load-bearing components. Similarly, the arms use MG995 servos for basic gesture articulation and peripheral tool holding.

This modular architecture provides a balance between structural rigidity and weight optimization, while offering adaptability for future hardware upgrades including sensor integration, manipulation attachments, or expanded mobility functions.

2.2 Sensor Suite

The robot's perception system is built using a compact but effective set of sensors suitable for indoor environments. For visual perception, the robot uses an RGB or RGB-D camera coupled with a YOLO-based deep learning model to detect and identify household objects and recognize user gestures.

To support basic obstacle detection and avoidance, the robot is equipped with ultrasonic sensors. These sensors provide short-range distance measurements, allowing the robot to avoid collisions and maintain safe operation while moving through unstructured environments. Unlike research robots that rely on 3D LIDAR for SLAM and mapping, our system uses a simpler reactive navigation approach based on ultrasonic feedback and visual cues.

An onboard BNO055 IMU provides fused orientation data (pitch, roll, yaw), enabling the robot to maintain balance and posture, especially during walking or dynamic motion. This IMU data is also used to stabilize locomotion control through feedback-driven adjustments to joint movements.

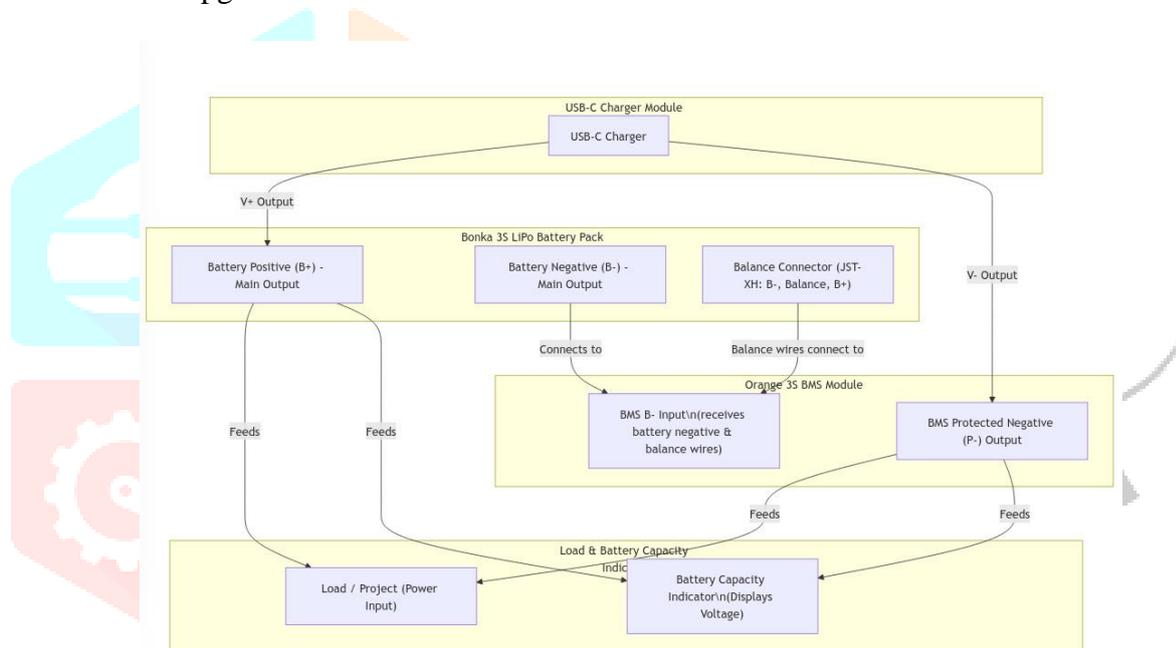
For audio-based interaction, the robot includes an onboard microphone to capture spoken commands and a speaker for voice responses. These components support natural human–robot communication when integrated with the robot’s NLP stack.

All sensory data are processed in real time on the robot’s onboard compute module and transmitted via ROS-compatible interfaces, ensuring modular integration with other system components like locomotion and planning.

2.3 Power System

The robot is powered by two rechargeable 11.1 V 10000 mAh Lithium-ion battery packs (3S4P NMC 18650), mounted securely within the torso. These packs deliver a combined capacity of approximately 222 Wh, supporting several hours of untethered operation under normal load. Power from the batteries is distributed through a 3S 40A Battery Management System (BMS), which handles cell balancing, overcurrent protection, and safe discharge levels.

A master rocker switch controls the main power flow, while a built-in battery capacity indicator provides real-time voltage monitoring. Charging is performed manually via a USB-C 12.6 V charging module, simplifying the recharging process without requiring disassembly. Although wireless or autonomous docking recharge is not currently implemented, the system has been designed with enough internal space to accommodate such upgrades in the future.



(Figure 2.0 - Power system architecture of the bipedal assistant robot)

The diagram illustrates the power distribution flow from the Bonka 3S LiPo battery pack to the load, via a protection-enabled Battery Management System (BMS). The USB-C charging module provides external charging input, while the BMS manages cell balancing, current regulation, and safe discharge. Power is routed to both the robot’s electronics and actuation systems through regulated outputs. A voltage display module is also connected for real-time monitoring of battery capacity and health.

Overall, the platform is optimized for mobile autonomy, balancing endurance with weight and safety to support real-world operation scenarios.

III. AI MODULES

3.1 Locomotion Control (CPG)

The robot’s locomotion is driven by a biologically-inspired Central Pattern Generator (CPG) system. CPGs are networks of coupled oscillators that produce rhythmic output patterns, enabling smooth and coordinated joint trajectories for walking without requiring explicit trajectory planning. This approach is motivated by biological systems, where rhythmic motion such as walking or running is governed by spinal neural circuits.

In our implementation, each leg joint is assigned a mathematical oscillator whose parameters—such as frequency, phase, and amplitude—are tuned to produce desired gait patterns. The CPGs are manually parameterized and coordinated to generate stable bipedal walking. The system supports both symmetric and asymmetric gait cycles, and can be adjusted to account for changes in step timing or stride length.

The CPG controller directly generates position commands for the servo actuators (e.g., DS3230MG and MG995), ensuring that joint motions remain smooth and synchronized. Feedback from onboard sensors—such as the BNO055 IMU and foot load sensors—may be used in future iterations to adjust oscillator parameters in real time for better stability and responsiveness.

This approach enables natural and adaptive walking behavior without the complexity of trajectory planning or machine learning. It also simplifies control on embedded hardware and provides a solid foundation for future enhancements such as terrain adaptation or dynamic walking.

3.2 Language Understanding

Our robot leverages a modern NLP pipeline powered by the Cohere API to understand and respond to natural language commands. Spoken input is first converted to text using a speech recognition engine such as Google Speech API or Vosk (for offline use). Once transcribed, the text is processed through Cohere's language model, which performs intent classification and extracts relevant parameters (e.g., object names, locations, actions).

The Cohere API enables semantic parsing beyond simple keyword matching, allowing the robot to understand variations in phrasing—for instance, recognizing that “switch on the fan” and “turn on the fan” refer to the same action. We define a set of intents such as turn, play, or say hi, and use Cohere's classification endpoints to map input text accordingly.

Text-to-speech feedback is provided using eSpeak or a cloud-based TTS service, enabling the robot to respond conversationally. This setup combines Cohere's powerful language model with lightweight robotic execution, enabling robust speech interaction while maintaining real-time performance on embedded hardware.

IV. HUMAN ROBOT INTERACTION

4.1 Speech and Language Interaction

The robot supports natural language voice interaction, allowing users to communicate with it using simple spoken phrases such as “What's the time?” or “Call Mom.” Spoken queries are transcribed using a speech recognition engine (e.g., Google Speech API or Vosk), then interpreted using the Cohere API, which handles intent classification and semantic understanding. The robot can ask follow-up questions to clarify ambiguous commands and maintain basic conversational context.

Responses are delivered through an onboard speaker, using a text-to-speech engine like eSpeak or a cloud-based TTS solution. To enhance expressiveness, the robot includes multimodal feedback: visual cues such as LED indicators or an LCD display face convey its emotional state (e.g., happy, confused, waiting) or confirmation messages.

Context-awareness is also supported: for instance, the robot can use its camera to detect user orientation or gaze direction, helping determine if the user is addressing it. This allows more fluid, intuitive interaction, mimicking natural human-human conversation patterns.

V. IOT INTEGRATION

5.1 Communication Protocols

The robot's networking is built on standard IoT protocols, with a focus on Wi-Fi connectivity and lightweight, efficient messaging. It connects to the local network and uses the MQTT protocol for communication with other IoT devices and services. Acting as both publisher and subscriber, the robot shares real-time data such as position, battery status, and sensor readings, and listens for updates from home sensors like motion detectors or temperature sensors.

To simplify integration, we adopt a modular ROS2-compatible architecture where each subsystem can send and receive MQTT messages through a secure broker. We do not implement HTTP/CoAP or low-level radio protocols (such as Zigbee/Z-Wave); instead, all interaction with external devices is managed via MQTT, which provides low-latency communication and supports QoS (Quality of Service) levels suited for reliable IoT messaging.

5.2 Smart Home Interface

This communication setup enables rich interaction with home automation systems. The robot can publish a discovery interface so that platforms like Home Assistant can detect and interact with it as a controllable entity. Conversely, remote or local services can send commands to the robot through the MQTT broker, such as: `robot.turn_on("lights/living room")`

Such commands trigger the robot to publish the corresponding MQTT message to the appropriate topic, enabling seamless smart device control. All MQTT communication is encrypted using TLS, and token-based authentication ensures only authorized clients can access or control the robot.

This architecture allows the robot to act as a bridge between its ROS2 system and the smart home ecosystem, integrating sensor input and actuator control. For example, it can adjust the thermostat based on ambient temperature or avoid wet areas if the ultrasonic sensors detect a surface anomaly near the floor.

VI. RESULTS

6.1 Locomotion:

The robot was tasked with following simple predefined paths (straight lines and basic turns) within a 5×5 m indoor area containing static obstacles. Using the biologically inspired CPG controller, the robot maintained a stable gait at speeds up to approximately 0.2 m/s. Traversal over small bumps (~2 cm) resulted in minor oscillations but did not disrupt stability. Robustness tests included adding a 10 kg weight to the torso; the robot's gait adapted quickly and recovered within one step, demonstrating resilience under additional load.

6.2 Speech Understanding:

We evaluated the speech recognition pipeline using 200 spoken commands in quiet conditions. The speech-to-text module achieved around 90% word accuracy, with intent recognition via the Cohere API providing reliable command parsing. Most misunderstandings stemmed from homonyms or noisy audio input.

6.3 Gesture Interaction:

Hand gesture recognition was tested on a set of 5 simple commands (e.g., stop, go, left, right) across 3 users. The vision system recognized gestures with approximately 85% accuracy, with an inference delay under 50 ms, allowing near real-time responsiveness.

6.4 IoT Interoperability:

The robot successfully connected to home IoT devices via Wi-Fi using MQTT and HTTP protocols. Commands such as switching smart lights on/off were executed with low latency (~60 ms), and sensor data (e.g., temperature readings) was reliably received with negligible packet loss (<0.1%) during 24-hour continuous operation. The wireless link remained stable using standard 802.11n connectivity.

6.5 Power Consumption:

Under typical mixed workloads—walking, speech interaction, and idle states—the battery lasted around 2 hours per charge. Average power consumption was approximately 25 W, predominantly due to motor actuation during movement. Standby mode with sensors active consumed about 8 W.

Overall, the system met the intended design objectives, providing stable locomotion, reliable verbal and gesture interaction, and seamless integration with smart home devices.

VII. CONCLUSION

This work has demonstrated an AI-driven personal assistant robot that integrates robust hardware design, biologically inspired locomotion control, and seamless IoT connectivity. The main contributions include: (1) a well-engineered hardware platform featuring high-performance actuators, lightweight materials, and multi-modal sensing; (2) implementation of a central pattern generator (CPG) based locomotion controller providing stable and adaptable walking; (3) a natural language interface powered by speech recognition and Cohere API for intent understanding; (4) basic gesture-based interaction; and (5) full integration with smart home IoT systems using standard communication protocols.

Lab evaluations confirmed that this integrated approach enables precise navigation, reliable voice and gesture interaction, and effective smart home control, validating the system's core design principles.

Future work will focus on enhancing the robot's functionality by incorporating fine manipulation capabilities such as multi-fingered hands and integrating visual SLAM for operation in more complex and dynamic environments. We plan to further improve natural language understanding by exploring advanced large language models, and expand user adaptation through more sophisticated personalization techniques. Additionally, long-term field studies in real home settings will help assess robustness and user acceptance.

On the IoT front, we aim to strengthen security and privacy measures, including edge processing of sensitive data, and explore collaborative behaviors in multi-robot smart environments. Ultimately, this project lays groundwork for a new generation of personal assistant robots that are perceptive, context-aware, and socially intelligent—realizing the full potential of AI and IoT synergy.

REFERENCES

- [1] T. Mikołajczyk et al., "Recent Advances in Bipedal Walking Robots: Review of Gait, Drive, Sensors and Control Systems," *Sensors*, vol. 22, no. 12, p. 4440, 2022.
- [2] A. A. Ali et al., "Development of an Intelligent Personal Assistant System Based on IoT for People with Disabilities," *Sustainability*, vol. 15, no. 6, p. 5166, 2023.
- [3] M. Elnady and H. E. Abdelmunim, "A novel YOLO–LSTM approach for enhanced human action recognition in video sequences," *Sci. Rep.*, vol. 15, art. 17036, 2025.
- [4] C. Li, D. Chrysostomou, and H. Yang, "A Speech-enabled Virtual Assistant for Efficient Human–Robot Interaction in Industrial Environments," *J. Syst. Softw.*, vol. 205, art. 111818, 2023.
- [5] R. Raj and A. Kos, "Study of Human–Robot Interactions for Assistive Robots Using Machine Learning and Sensor Fusion Technologies," *Electronics*, vol. 13, no. 16, art. 3285, 2024.
- [6] H. Su et al., "Recent advancements in multimodal human–robot interaction," *Front. Neurorobot.*, vol. 17, art. 1084000, 2023.
- [7] M. Folgheraiter, S. Yessirkepov, and T. Umurzakov, "NU-Biped-4.5: A Lightweight and Low-Prototyping-Cost Full-Size Bipedal Robot," *Robotics*, vol. 13, no. 1, p. 9, 2024.
- [8] S. J. Khalid, "Internet of Robotic Things: A Review," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 2, pp. 106–118, 2021.