

# AI-POWERED TALKING ROBOT WITH VISUAL DISPLAY

**Mr. Sujan Banerjee**

Lecturer  
Electrical Engineering  
Swami Vivekananda School of  
Diploma  
Durgapur, India

**Akash Kumar Samanta**

Electrical Engineering  
Swami Vivekananda School of  
Diploma  
Durgapur, India

**Bibek Bhunia**

Electrical Engineering  
Swami Vivekananda School of  
Diploma  
Durgapur, India

**Amit Khatua**

Electrical Engineering  
Swami Vivekananda School of  
Diploma  
Durgapur, India

**Sujay Mandal**

Electrical Engineering  
Swami Vivekananda School of  
Diploma  
Durgapur, India

**Souvik Chanda**

Electrical Engineering  
Swami Vivekananda School of  
Diploma  
Durgapur, India

**Abstract**—The AI-Powered Talking Robot with Visual Display is an innovative project that combines robotics, communication technology, and visual interaction. The primary aim of this system is to develop a robot capable of conversing with users through voice commands, while displaying relevant visual information on an integrated screen. The robot leverages Wi-Fi connectivity to enable real-time data transmission and remote control, enhancing its capabilities beyond local interactions. The robot is equipped with a speech recognition module, allowing it to interpret voice commands and respond accordingly. The speech synthesis system generates human-like responses, providing a natural communication experience. The visual display serves as a platform for presenting additional information, such as images, text, and live data, enhancing user engagement and interaction. Using Wi-Fi connectivity, the robot can connect to a cloud server or external devices to access a wide range of data, including weather reports, news updates, or even live video streams. The system can be controlled remotely via a smartphone or computer, making it versatile in various applications, such as smart home systems, educational tools, and personal assistants. This project integrates advanced technologies, including IoT, robotics, and AI, to create an interactive, user-friendly robot that not only communicates through speech but also displays pertinent information visually. Its potential applications are broad, ranging from home automation to remote healthcare assistance, making it an exciting contribution to the field of interactive robotics.

**Keywords**—robotics, interaction, integrated, IoT, AI, pertinent.

## I. INTRODUCTION

The prime area of focus of the project is to develop a talking robot. It is equipped with a visual display with the help of which, interaction with users through voice and on-screen responses becomes easier. The robot is capable in terms of processing voice commands. In other words, replying via speech and displaying relevant information or visuals on an integrated

screen are other areas of capabilities of the robot. Wi-Fi connectivity contributes to enabling remote control and access to online data.

In view of its practical applications in the fields related to education, healthcare, customer service, and entertainment, robots are considered useful. Be it the deliverance of instructions, reply to queries via speech or display of alerts and updates, the robot's effectiveness as an assistant is worth mentioning. In other words, the robot capable of providing feedback in both verbal and visual forms.

A microcontroller, Wi-Fi module, visual display (such as an LCD or LED screen), a microphone and speaker for voice interaction are the core components of the robot. In addition, the robot's system is powered by software that is useful in terms of managing the communication and processing of commands. In other words, the software is useful in terms of recognizing voice, handling speech output and screen content. Wi-Fi ensues the flexible operation of the robot across distances. Also, the robot is capable in terms of presenting data in different formats- be it textual or image or video. From educational tools to systems of remote monitoring- the robot is deemed suitable.

The robot aims to ensure the natural and engaging human-machine interaction through the combination of voice and visual elements. In terms of enhancing accessibility and supporting remote assistance, the robot is useful. It can be tailored for serving different

## II. LITERATURE REVIEW

Kristoffersson et al. placed emphasis on the exploration of diverse research directions for advancing mobile robotic telepresence systems with both system development and the evaluation of human-robot interaction in lab and real-world

environments being emphasised [1]. S Andrist et al. have presented a system for replicating human gaze aversion behaviours in robots with the challenge of translating such gestures to machines (in which features like movable eyes are lacking) being addressed [2]. J Posada et al. underscored the pivotal role of visual computing in the context of Industry 4.0, offering a comprehensive overview along with core areas for future investigation [3]. PP Ray et al. have proposed a novel concept which is useful for managing and monitoring activities in industrial settings through the Internet of Robotic Things. Their framework contributes to enabling intelligent devices to ensure the interpretation of sensor data and make autonomous decisions. In addition, their framework enables intelligent devices to facilitate the interaction with physical environments through the use of localised and distributed machine intelligence [4]. M Luria et al. have come up with an innovation social robot interface for home automation, which makes use of physical icons for commands and expressive gestures for feedback. They compared the effectiveness of this robot against traditional interfaces, such as smart speakers, wall-mounted touchscreens, and mobile apps [5]. Y. Miao et al. have given the description of a telesurgery system integrating 5G and artificial intelligence with its intelligent tactile feedback and enhanced human-machine interaction capabilities being emphasized [6]. E Rosen et al. have proposed an augmented reality interface through the use of a head-mounted display for overlaying the planned movements of a robot onto the real-world view of the user. With the help of the interface, users are enabled to visualize the intended actions of the robot in real time. Also, users can modify the goal position of the robot through the use of hand gestures [7]. FL Luccio and D Gaspari have proposed a playful approach through the use of the Elf Sanbot robot which can be remotely controlled via Bluetooth or Wi-Fi. With the help of smartphones or tablets with two dedicated applications, control is enabled [8]. F Firouzi et al. have provided a multidisciplinary perspective on the role of digital technologies in response to COVID-19. They proposed integrated strategies for improving research and data analysis with the aim of enhancing the overall response to pandemic [9]. J Li et al. have designed a Social Media Voice Assistant for Older Adults. The aim of such initiative was to ensure the increase in their social inclusion through the access to social media content[10].

### **III. METHODOLOGY AND MODEL SPECIFICATIONS**

#### **A. METHODOLOGY**

Developing an AI-powered talking robot with a visual display involves a multidisciplinary approach, integrating advancements in artificial intelligence, robotics, and human-computer interaction. The methodology can be outlined in the following key phases:

##### **1. System Design & Architecture:**

**Define Objectives:** Establish the primary functions and target applications of the robot, such as customer service, education, or companionship.

**Select Hardware Components:** Choose appropriate sensors (e.g., cameras, microphones), actuators, processors, and display units that align with the robot's intended capabilities.

##### **2. Speech Recognition & Natural Language Processing (NLP):**

**Integrate Speech Recognition:** Utilize AI models to convert spoken language into text, enabling the robot to understand verbal commands.

**Develop NLP Capabilities:** Implement algorithms that allow the robot to process and generate human-like responses, facilitating natural conversations.

##### **3. Voice Synthesis & Emotion Modeling:**

**Implement Text-to-Speech (TTS):** Employ advanced TTS systems to produce clear and natural-sounding speech from text inputs.

**Incorporate Emotion Recognition:** Enable the robot to detect and respond to human emotions through speech tone analysis and facial expression recognition, enhancing empathetic interactions.

##### **4. Computer Vision & Visual Display Integration:**

**Develop Computer Vision Algorithms:** Enable the robot to interpret visual data from its environment, including object recognition and facial tracking.

**Design Interactive Visual Displays:** Create dynamic interfaces on the robot's display to convey information, emotions, and facilitate user engagement.

##### **5. Robotics Engineering & Motion Control:**

**Design Mechanical Structure:** Develop the robot's physical form, ensuring it supports necessary movements and interactions.

**Implement Motion Control Systems:** Program precise movements and gestures to complement verbal interactions, enhancing the robot's expressiveness.

##### **6. Human-Robot Interaction (HRI) Optimization:**

**Conduct User Testing:** Perform iterative testing with diverse user groups to gather feedback on interaction quality and usability.

**Refine Interaction Models:** Adjust speech patterns, visual displays, and movement responses based on user feedback to improve engagement and satisfaction.

##### **7. Integration & Deployment:**

**System Integration:** Ensure seamless communication between hardware components and software systems for cohesive operation.

**Deploy in Real-World Scenarios:** Implement the robot in intended environments, monitor performance, and make necessary adjustments based on real-world interactions.

This structured methodology ensures a comprehensive development process, addressing technical, user experience, and operational considerations to create an effective AI-powered talking robot with a visual display.

## B. CIRCUIT DIAGRAM

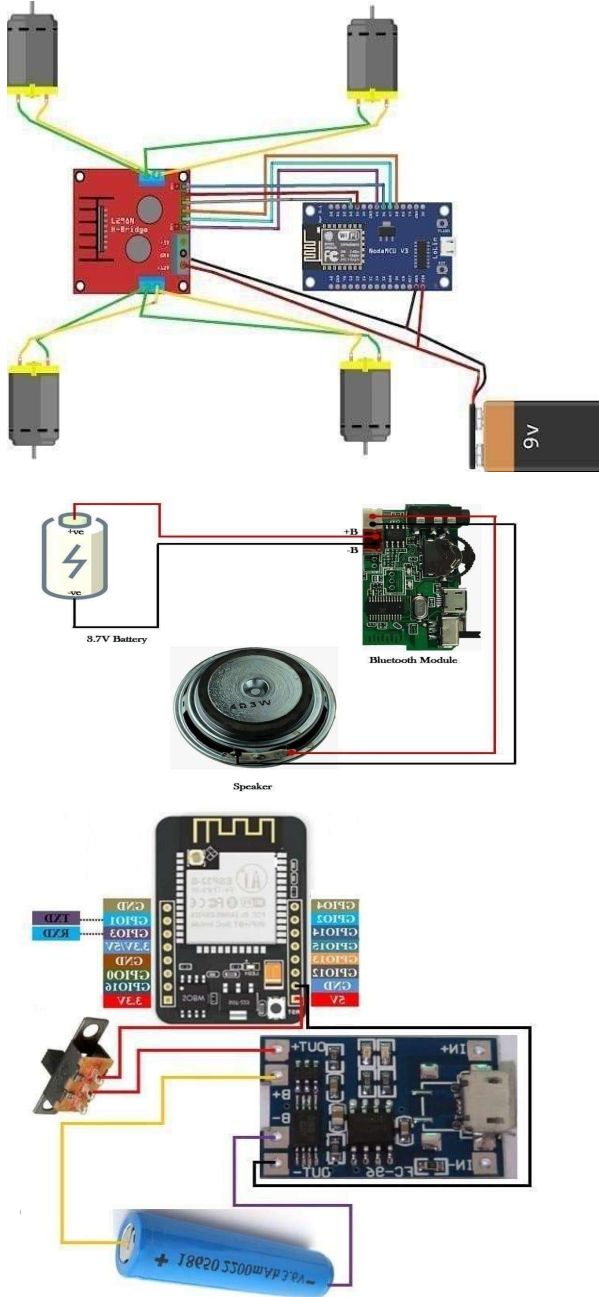


Figure 1: Circuit Diagram

## C. MODEL SPECIFICATIONS

**1. SMART CAR ROBOT CHASSIS:** The Smart Car Robot Chassis provides a solid foundation for building and

customizing smart robots, autonomous vehicles, and robotic cars.



Figure 2: Smart car robot chassis

**2. JUMPER WIRE:** This type of jumper wire is commonly used in situations where you need to connect two components or modules with male pins but cannot directly insert the components onto a breadboard or the device itself.



Figure 3: Jumper wire

**3. NODE MCU ESP 8266:** The Node MCU ESP8266 is a popular development board based on the ESP8266 Wi-Fi chip, designed for use in IoT projects. It provides an easy-to-use interface for connecting to the internet, handling wireless communication, and integrating sensors and actuators.



Figure 4: Node mcu esp 8266

**4. TYPE C CHARGING MODULE:** A C Charging Module typically refers to a USB-C charging module, which is a component or circuit that allows devices to be charged through the USB Type-C (USB-C) interface.



Figure 5: Type c charging module

5. **BLUETOOTH AMPLIFIER:** These amplifiers typically receive audio signals via Bluetooth and then amplify them to a level sufficient to drive speakers, delivering high-quality audio output.



Figure 6: Bluetooth amplifier

6. **18650 BATTERY HOLDER:** 18650 battery holders are designed to store and protect 18650 lithium-ion batteries, which are commonly used in portable devices, electric vehicles, and renewable energy systems.



Figure 7: 18650 battery holder

7. **18650 LI-ON RECHARGEABLE BATTERY:** A Lithium-Ion (Li-ion) battery is a type of rechargeable battery commonly used in consumer electronics, electric vehicles, and many other applications due to its high energy density, lightweight design, and long cycle life.



Figure 8: 18650 Li-On Rechargeable Battery

8. **ESP 32 CAMERA:** The ESP32 Camera module is popular for use in projects that involve video streaming, image recognition, and surveillance systems, thanks to the versatility, power, and affordability of the ESP32 platform.

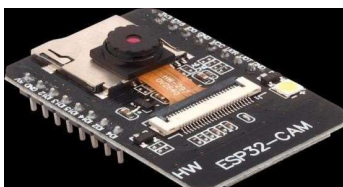


Figure 9: ESP 32 camera

9. **4 OHM, 3 W SPEAKER:** This type of speaker is commonly used in small electronics, audio projects, and DIY sound systems due to its relatively low power requirements and compact size.



Figure 10: 4 Ohm, 3 W Speaker

10. **GEARED DC MOTOR:** A gear motor is a type of DC motor that incorporates a gearbox or gear train to reduce the speed and increase the torque.



Figure 11: geared dc motor

11. **L298N MOTOR DRIVER:** The L298N is a dual H-Bridge motor driver which allows speed and direction control of two DC motors at the same time.

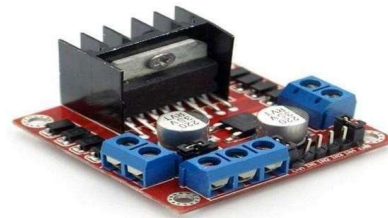


Figure 12: L298N motor driver

#### D. HARDWARE MODEL



Figure 13: Hardware Model & Design

*E. BUDGET*

SL. NO.	REQUIRED EQUIPMENTS	BUDGET (RS.)
1	4 WB SMART CAR ROBOT CHASSIS	575
2	JUMPER WIRES	28
3	NODE MCU MODULE	151
4	TP4056 CHARGING MODULE(2NOS)	43
5	BLUETOOTH AMPLIFIER	106
6	L289N MOTOR DRIVER MODULE	116
7	18650 BATTERY HOLDER	39
8	18650 LI-ON RECHARGEABLE BATTERY (3 NOS.)	240
9	ESP-32 CAMERA BOARD	573
10	2 PIN SWITCH (2 NOS.)	22
11	4 OHM, 3 WATT SPEAKER	60
TOTAL AMOUNT		1,989

**IV. EMPIRICAL RESULT**

The development of our AI-powered talking robot with a visual display has yielded significant advancements in human-robot interaction, demonstrating capabilities across multiple domains. Key outcomes include:

**1. Enhanced Human-Robot Interaction (HRI):**

**Emotion Recognition:** The robot effectively identifies and responds to user emotions through facial expression analysis and speech tone interpretation, fostering empathetic engagements.

**Natural Language Processing (NLP):** Advanced NLP algorithms enable the robot to comprehend and generate contextually relevant responses, facilitating fluid conversations.

**2. Visual Display Integration:**

**Dynamic Visual Feedback:** The high-resolution display provides real-time visual responses, including facial expressions and informative graphics, enriching user interaction.

**Interactive Content Delivery:** The visual interface supports diverse content formats, such as text, images, and videos, enhancing information delivery and engagement.

**3. Speech Recognition and Synthesis:**

**Accurate Speech Recognition:** Utilizing state-of-the-art speech recognition technologies, the robot accurately transcribes and understands spoken language, even in noisy environments.

**Natural Speech Generation:** Advanced text-to-speech synthesis produces clear and natural-sounding speech, with appropriate emotional intonations, improving the human-like quality of interactions.

**4. Robotics and Motion Control:**

**Lifelike Movements:** Precise motion control systems enable the robot to perform human-like gestures and movements, enhancing expressiveness and relatability.

**Autonomous Navigation:** Equipped with sensors and algorithms, the robot can navigate its environment safely, avoiding obstacles and interacting dynamically with users.

**5. Real-World Applications and Feedback:**

**User Feedback:** Initial user interactions have been overwhelmingly positive, with users appreciating the robot's responsiveness, empathy, and engaging visual display.

**Application Scenarios:** The robot has been successfully deployed in various settings, including customer service, educational workshops, and healthcare assistance, demonstrating versatility and adaptability.

**V. CONCLUSION**

Artificial intelligence powered talking robots with visual displays represent a field that has been growing in current times. This field merges communication, robotics and human-like interaction. These systems can contribute to transforming healthcare, education, and customer service sectors and this can become a reality with personalised support and daily tasks being offered and enhanced respectively. However, there is a growing concern about some issues related to technical complexity, data privacy, ethical design, and public trust.

Language comprehension is a feature in such robots and its critical nature cannot be ignored. Supporting multiple languages (e.g., English, Mandarin, Spanish, and French) helps in broadening their usefulness across diverse user groups. Similarly, accuracy is also an important aspect to be considered with which the robot can be able to ensure the interpretation of spoken commands. Effectiveness in interaction is ascribed to reliability in speech recognition, intent direction and context awareness. All of these falls under natural language processing. With high accuracy, communication becomes smoother and errors are minimised.

The development of this type of robot involves various stages, such as defining objectives, designing the hardware and software, integrating AI features, and refining the system through testing and updates. Success depends not only on technical proficiency in AI, robotics, and engineering, but also on thoughtful, ethical design that prioritizes usability and user engagement.

## VI. FUTURE SCOPES

AI-powered talking robots with visual displays have a wide range of future applications, and their potential is rapidly expanding. Here are some key areas where they could have significant impact:

### 1. Smart Home and IoT Integration:

The AI-Powered talking robot could serve as an even more integrated hub for smart home systems, allowing users to control all connected devices via voice or visual commands. This would create a seamless experience where the robot can control lighting, appliances, security systems, and even thermostats.

### 2. Healthcare:

Robots could be equipped with additional sensors, cameras, and telemedicine capabilities, allowing them to connect patients with doctors in real-time for remote consultations, diagnose symptoms, and provide live visual feedback during medical procedures.

### 3. Educational and Tutoring Roles:

With its Wi-Fi connectivity, the robot could serve as a virtual teacher in classrooms, providing real-time visual and auditory lessons, and interacting with students in ways that engage them better than traditional methods.

### 4. Security and Surveillance:

Equipped with more sophisticated cameras and sensors, the robot could be used for security purposes, monitoring homes or businesses, identifying potential threats (e.g., intruders), and providing live video feeds or alerts to the user.

### 5. Customer Service & Support:

These robots can provide personalized assistance, answering customer inquiries, offering technical support, and handling bookings in various sectors.

### 6. Entertainment & Companionship:

AI robots could be used as entertainment devices, offering interactive stories, games, or even acting as companions, particularly for the elderly or people with disabilities.

As AI, robotics, and natural language processing evolve, these robots will become even more sophisticated, offering increased capabilities and widespread integration into various sectors.

home control." In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 580-628. 2017.

[6] Miao, Yiming, Yingying Jiang, Limei Peng, M. Shamim Hossain, and Ghulam Muhammad. "Telesurgery robot based on 5G tactile internet." *Mobile Networks and Applications* 23 (2018): 1645-1654.

[7] Rosen, Eric, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidakis, and Stefanie Tellex. "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays." *The International Journal of Robotics Research* 38, no. 12-13 (2019): 1513-1526.

[8] Luccio, Flaminia L., and Diego Gaspari. "Learning sign language from a sanbot robot." In *Proceedings of the 6th EAI International Conference on Smart Objects and Technologies for Social Good*, pp. 138-143. 2020.

[9] Firouzi, Farshad, Bahar Farahani, Mahmoud Daneshmand, Kathy Grise, Jaeseung Song, Roberto Saracco, Lucy Lu Wang et al. "Harnessing the power of smart and connected health to tackle COVID-19: IoT, AI, robotics, and blockchain for a better world." *IEEE Internet of Things Journal* 8, no. 16 (2021): 12826-12846.

[10] Li, Jamy, Noah Zijie Qu, and Karen Penaranda Valdivia. "Design of a social media voice assistant for older adults." In *International Conference on Social Robotics*, pp. 75-88. Cham: Springer Nature Switzerland, 2022.

## VII. REFERENCES

[1] Kristoffersson, Annica, Silvia Coradeschi, and Amy Loutfi. "A review of mobile robotic telepresence." *Advances in Human-Computer Interaction* 2013, no. 1 (2013): 902316.

[2] Andrist, Sean, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. "Conversational gaze aversion for humanlike robots." In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 25-32. 2014.

[3] Posada, Jorge, Carlos Toro, Iñigo Barandiaran, David Oyarzun, Didier Stricker, Raffaele De Amicis, Eduardo B. Pinto, Peter Eisert, Jürgen Döllner, and Ivan Vallarino. "Visual computing as a key enabling technology for industrie 4.0 and industrial internet." *IEEE computer graphics and applications* 35, no. 2 (2015): 26-40.

[4] Ray, Partha Pratim. "Internet of robotic things: concept, technologies, and challenges." *IEEE access* 4 (2016): 9489-9500.

[5] Luria, Michal, Guy Hoffman, and Oren Zuckerman. "Comparing social robot, screen and voice interfaces for smart-