IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15381

3D HUMANOID AI-SMART "AW8R"

Prof. CHANDRESH BHANGE[1], PRATYUSH.G. JANBHANDHU[2], SAMIKSHA A. GAJBHIYE[3], PRAKRUTI R. GAIMUKHE[4], TANU R. PATALE[5], RAVISHANKAR V. SHAHU[6], JAYASHRI M. MARBATE[7]

lecturer at Department of Electronics & Telecommunications Engineering in NIT Polytechnic, Nagpur[1]

student at Department of Electronics & Telecommunications Engineering at NIT Polytechnic, Nagpur[2-7]

Abstract: The 3D Humanoid Al-Smart "AW8R" is an advanced portable Al-based voice assistant system designed to enhance natural human-machine interaction by integrating speech recognition, emotion detection, and a 3D holographic humanoid avatar. Unlike traditional voice assistants that rely solely on audio responses, the proposed system introduces synchronized facial expressions and lip movements to create a more engaging and realistic user experience.

The system is built around the ESP32-S3 microcontroller with integrated Wi-Fi for cloud connectivity. Audio input is captured using an INMP441 MEMS microphone and transmitted to cloud-based Al services such as Gemini or Open Al for speech-to-text conversion, intent detection, and contextual response generation. The backend also performs emotion analysis and text-to-speech synthesis. The processed response is delivered to the user through a 3W speaker driven by a MAX98357A amplifier, while a 1.3-inch TFT display renders a 3D humanoid avatar. A holographic acrylic pyramid creates a floating 3D visual effect.

The system is powered by a 2400mAh Li-ion battery with integrated charging and boost regulation. Experimental evaluation demonstrates stable operation, synchronized animation, and improved user engagement compared to traditional voice-only systems. The project provides a cost-effective and scalable solution for next-generation Al assistants.

I. INTRODUCTION

Artificial Intelligence-based voice assistants have become widely adopted in smart homes, offices, and digital ecosystems. Systems such as Alexa and Google Assistant use natural language processing (NLP) to interpret spoken commands and provide appropriate responses. However, most commercial systems lack expressive visual feedback, limiting the naturalness of interaction.

The 3D Humanoid Al-Smart AW8R project introduces an emotionally expressive humanoid avatar integrated with Al processing to improve user engagement and communication realism.

II. LITERATURE SURVEY

1. Previous Al assistants focus primarily on:
Voice recognition
NLP response generation
Smart device control
Research in embodied Al suggests that adding visual emotional cues significantly improves user experience. Some studies implemented robotic faces or 2D animated avatars, but real-time lip synchronization and compact embedded integration remain limited.

2. Lack of 3D humanoid avatars
Poor emotion synchronization
High power consumption
Limited portability
Audio-visual delay mismatch

3. Literature indicates the need for emotionally intelligent and visually expressive Al systems. The proposed AW8R system addresses these gaps by integrating cloud Al processing with embedded 3D avatar rendering and holographic

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15381

visualization.

III. SYSTEM ARCHITECTURE

The proposed system is a portable humanoid Al assistant device designed using the ESP32-S3 microcontroller. The system integrates audio input, Al processing, animated visual output, and holographic projection to create a human-like interactive assistant.

The architecture consists of several interconnected modules:

1. Input Unit

INMP441 Digital Microphone captures the user's voice commands.

2. Processing Unit

ESP32-S3 microcontroller processes voice data and controls the entire system.

3. Visual Output Unit

ST7789 1.3-inch TFT display renders a humanoid Al face with animations such as eye blinking and lip synchronization.

4. Audio Output Unit

MAX98357A digital audio amplifier drives a 40 speaker to produce audio responses.

5. Power Management Unit

TP4056 lithium battery charger module manages battery charging.

MT3608 boosts converters regulate the voltage supply to power the system.

6. Projection System

A transparent pyramid placed above the display creates a holographic illusion of the Al assistant.

Data Flow
User Voice Microphone ESP32
Processing
Audio Output via Speaker + Visual Output
via TFT Display
Holographic Projection

IV. METHOD & MATERIAL

The system was developed using the embedded system design methodology which involves the following steps:

1. Selection of suitable hardware components.

2. Design of circuit schematic using PCB design software.

3. PCB layout design and fabrication.

4. Firmware development using ESP32 programming environment.

5. System integration and testing.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15381

The design uses digital communication protocols such as:

SPI (Serial Peripheral Interface) for the TFT display

12S (Inter-IC Sound) for microphone and amplifier communication.

Materials Used

Component Description

ESP32-S3 Dev Board Main microcontroller

INMP441 MEMS Microphone Digital audio input

MAX98357A Amplifier12S audio amplifier

ST7789 TFT Display 1.3-inch graphical display

TP4056 Module Lithium battery charging module

MT3608 Module DC-DC boost converter

2400mAh Li-ion Battery Portable power source

40 Speaker Audio output

Toggle Switch Power control

Custom PCB Circuit integration board

V. HARDWARE DESCRIPTION

ESP32-S3 Microcontroller

The ESP32-S3 microcontroller acts as the central processing unit of the system. It processes audio input, manages peripheral communication, and controls display animations.

Features:

Dual-core processor
Wi-Fi and Bluetooth connectivity
Integrated digital interfaces (SPI, I2S)
Low power consumption
INMP441 Digital Microphone

The INMP441 is a MEMS digital microphone that communicates using the 12S protocol. It directly sends digital audio data to the ESP32 without requiring an analog-to-digital converter.

Key features:

High signal-to-noise ratio
Digital output
Low power consumption
Compact size
MAX98357A Audio Amplifier

The MAX98357A is a class-D digital audio amplifier that converts digital audio signals from ESP32 into analog signals suitable for driving a speaker.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15381

Key features:

12S digital audio interface
Up to 3W output power
High efficiency
Minimal external components
ST7789 TFT Display

The ST7789 display is a high-resolution color TFT display used to show the humanoid Al assistant interface.
Specifications:

Size: 1.3 inch

Resolution: 240 x 240 pixels

Interface: SPI

Color support: 65K colors

The display shows a humanoid Al face with animation, including:

eye blinking

lip movement during speech

idle expressions

Power Management System

The power system consists of two modules.

TP4056 Battery Charging Module

This module safely charges the lithium battery and provides protection against overcharging and over-discharge.
MT3608 Boost Converter

The MT3608 boosts the battery voltage to provide stable power for the ESP32 and other modules.

VI. WORKING METHODOLOGY

The system operates through the following sequence:

1. When the device is powered on, the ESP32 initializes all connected modules.

2. The TFT display loads the humanoid Al interface and displays the assistant in idle mode.

3. The INMP441 microphone continuously listens for voice input.

4. When a user speaks, the microphone captures the voice signal and sends it to the ESP32.

5. The ESP32 processes the audio input and generates a response.

6. During the response:

The speaker produces audio output.

The TFT display animates the Al face with lip synchronization.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

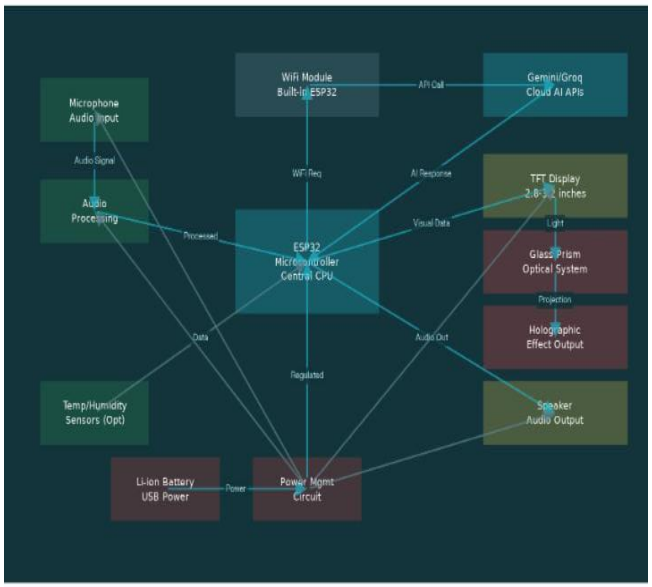
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15381

7. The Al avatar periodically performs eye blinking animation to simulate natural behavior.
8. The holographic pyramid reflects the display image to produce a 3D floating effect.

VII. BLOCK DIAGRAM OF 3D Humanoid AI-SMART AW8R



VIII. ADVANTAGES

The proposed system provides several advantages:

1. Portable battery powered design

2. Low power consumption

3. Human-like Al interface

4. Real-time voice interaction

5. Compact embedded system architecture

6. Cost-effective hardware components

7. Expandable system for future Al integration

IX. APPLICATION

This system can be used in various applications such as:
1. Smart home assistant devices
2. Educational learning assistants
3. Interactive Al companions
4. IOT device control interfaces
5. Exhibition and demonstration projects
6. Human-machine interaction research
7. Robotics and Al development platforms

X. RESULTS AND DISCUSSION

The developed prototype successfully demonstrated the integration of voice input, visual output, and audio feedback. Key observations:

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026
DOI: 10.17148/IJARCCE.2026.15381

The microphone accurately captured user voice commands.
The ESP32 efficiently processed audio data.
The TFT display rendered smooth animations for the humanoid Al face.
Eye blinking and lip synchronization improved the realism of the Al assistant.
The amplifier provided clear audio output through the speaker.
The holographic pyramid enhanced the user experience by creating the illusion of a floating Al assistant.
Overall, the system demonstrated stable operation and effective human-machine interaction.

XI. CONCLUSION

This project successfully demonstrates the design and implementation of a portable humanoid Al assistant system using ESP32-S3.
The integration of voice input, animated display output, and holographic projection creates a realistic Al interaction platform. The system provides a cost-effective and compact solution for developing embedded Al assistants.
Future improvements may include:
advanced Al speech recognition
cloud-based Al processing
emotion-based facial expressions
gesture recognition
improved holographic projection technology

REFERENCES

[1]. EEPRESSIF SYSTEMS - ESP32-S3 Technical Reference Manual
[2]. TDK INVENSE INMP441 MEMS Microphone Datasheet
[3]. Maxim Integrated - MAX98357A Audio Amplifier Datasheet
[4]. SITRONIX Technology-ST7789 TFT Display Controller Datasheet
[5]. EASYEDA-PCB Design Documentation
[6]. Arduino ESP32 Development Guide
[7]. 7.IEEE Embedded Systems Research Publication

.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 3, March 2026

DOI: 10.17148/IJARCCE.2026.15381

