MARVIN 3.0: AN INTELLIGENT EMOTIONAL PET ROBOT FOR ENHANCED HUMAN–ROBOT INTERACTION

Chandana J Nayak, Dr. Balapradeep K N

KVG COLLEGE OF ENGINEERING

Abstract: Human–Robot Interaction (HRI) has evolved significantly with advances in artificial intelligence, computer vision, natural language processing, and Internet of Things (IoT) technologies. This paper presents MARVIN 3.0 (Multi-faceted Adaptive Robotic Virtual Intelligence Nucleus), an intelligent emotional pet robot designed to provide personalized and emotionally engaging interactions. The proposed system introduces three major enhancements over previous versions: an emotion display screen, enhanced face recognition, and an improved wake-word detection pipeline. Built using ESP32-S3, Raspberry Pi, MTCNN-based face detection, Mobile Face recognition, and CNN-MFCC wake-word classification, the system enables real-time emotional expression and context-aware responses. Experimental targets indicate wake-word detection accuracy exceeding 95%, face recognition accuracy up to 97%, and emotion display response times below 500 ms. MARVIN 3.0 demonstrates the potential of emotionally intelligent robotic companions for home assistance, entertainment, and social interaction.

Keywords— Human–Robot Interaction, Emotion Recognition, Face Recognition, Wake-Word Detection, IoT, ESP32-S3, Raspberry Pi, Mobile Face, MTCNN.

I. INTRODUCTION

II. Related Work

Hasan et al. proposed MARVIN, an intelligent pet robot combining IoT, machine learning, computer vision, and NLP. The system achieved high wake-word and face recognition accuracy but lacked emotional display capabilities.

Badri Narayanan and Chandira Prabha developed an IoT-enabled robotic pet care system utilizing ESP32-CAM and voice-assisted automation. While effective for monitoring and control, the system lacked advanced AI-driven emotional interaction.

Gaitan-Padilla et al. introduced touch-sensitive social robot interactions using Fiber Bragg Grating sensors and machine learning classifiers. Although highly accurate, the approach requires specialized sensing hardware.

These studies demonstrate the growing importance of emotional intelligence and personalized interaction in robotic systems while highlighting opportunities for improvement in expressive communication.

III. Problem Statement

Current pet robots and smart assistants often provide mechanical responses without visible emotional feedback. Such systems struggle to maintain long-term user engagement and lack context-aware personalization.

Additionally, wake-word detection systems frequently suffer from noise interference, resulting in false activations or missed commands. Existing systems also rarely combine emotion display with user recognition, limiting the overall quality of interaction.

Therefore, there is a need for an intelligent robotic companion capable of emotional expression, robust voice activation, and personalized interaction.

IV. Proposed Methodology

A. System Architecture

MARVIN 3.0 adopts a three-layer architecture:

1. **Perception Layer** – Captures audio and visual inputs using ESP32-S3 and ESP32-CAM.
2. **Processing Layer** – Performs face recognition, intent classification, and emotion mapping on Raspberry Pi.
3. **Actuation Layer** – Controls emotion display, speaker output, and robot movement.

B. Wake-Word Detection

Audio signals are acquired through an I2S microphone at 16 kHz. Voice Activity Detection (VAD) removes background noise. MFCC features are extracted and processed using a CNN classifier trained to detect the wake phrase “Hey Marvin”.

C. Face Detection and Recognition

The ESP32-CAM captures video frames that are analyzed using MTCNN for face detection. Mobile Face generates facial embeddings which are compared against registered user profiles using cosine similarity.

D. Emotion Display System

An OLED/TFT display presents animated facial expressions such as happy, sad, angry, surprised, and neutral. Emotion states are dynamically selected based on conversational context and recognized user identity. The Emotion Display System is a core component of MARVIN 3.0 that enables the robot to communicate its emotional state to users in a natural and engaging manner. The system is designed to enhance human–robot interaction by providing visual, auditory, and behavioral cues that reflect the robot's perceived emotions. MARVIN 3.0 utilizes data obtained from sensors, facial expression recognition, voice analysis, and user interaction patterns to determine the emotional context of the interaction. Based on this information, the robot classifies emotions into predefined categories such as happiness, sadness, excitement, surprise, relaxation, and concern. The detected emotional state is displayed through an integrated LED display that changes facial expressions and eye patterns according to the robot's mood. Additionally, servo motors control body movements, including head tilts, tail wagging, and posture adjustments, to provide expressive non-verbal communication. Voice synthesis modules further enhance emotional expression by modulating speech tone, pitch, and response style.

E. IoT and Conversational Intelligence

Firebase Realtime Database enables smart home automation, while Gemini API integration provides intelligent conversational responses and multi-turn interaction. The Internet of Things (IoT) plays a crucial role in enhancing the functionality and connectivity of MARVIN 3.0. Through IoT technology, the robot can communicate with cloud services, smart devices, and remote monitoring platforms, enabling intelligent and seamless interactions. The system utilizes Wi-Fi and Bluetooth communication modules to exchange data between the robot, mobile applications, and cloud databases.IoT integration allows MARVIN 3.0 to store user preferences, interaction history, and emotional response patterns for personalized experiences. The cloud-based architecture facilitates real-time data processing, software updates, and remote access to robot status information. Furthermore, the robot can interact with smart home devices such as lights, speakers, and security systems to provide context-aware assistance and companionship.

V. Expected Results and Performance Evaluation

Metric	Target Value
Wake-word Detection Accuracy	>95%
Face Recognition Accuracy	Up to 97%
Emotion Display Response Time	<500 ms
IoT Command Latency	<2 seconds

The proposed enhancements are expected to improve user engagement, emotional responsiveness, and interaction reliability compared to conventional robotic assistants.

Conclusion

MARVIN 3.0 introduces a comprehensive framework for emotionally intelligent robotic companionship. By integrating emotion display, face recognition, wake-word detection, and IoT connectivity, the system aims to create more natural and meaningful human–robot interactions. Future work will focus on advanced emotion recognition, multimodal learning, and adaptive behavioral modeling to further enhance personalization and social engagement. This paper presented MARVIN 3.0: An Intelligent Emotional Pet Robot for Enhanced Human–Robot Interaction, a next-generation companion robot designed to provide emotionally aware and personalized interactions with users. By integrating emotion recognition, conversational intelligence, adaptive behavioral responses, and IoT connectivity, MARVIN 3.0 offers a more natural and engaging human–robot interaction experience compared to conventional robotic companions. The proposed system is capable of recognizing user emotions through facial expressions, voice analysis, and interaction patterns, enabling it to respond with appropriate visual, auditory, and behavioural cues. The incorporation of conversational intelligence allows meaningful communication, while IoT integration enhances connectivity, data management, and remote accessibility. These features collectively contribute to increased user engagement, emotional support, and companionship.

References

[1] M. R. Hasan, S. Saha, and M. M. Hossain, “MARVIN an Intelligent Pet Robot: A Convergence of IoT, Machine Learning, Computer Vision and NLP,” ICCIT, 2024.

[2] B. Narayanan and C. P. S, “Robotic Pet Care System Powered by IoT,” ICICNIS, 2024.

[3] M. Gaitan-Padilla et al., "Touch-Sensitive Hand Interactions for Social Robots Using Fiber Bragg Grating Sensors," IROS, 2025.

[4] A. Mehrabian, "Communication Without Words," Psychology Today, vol. 2, no. 4, pp. 53–56, 1968.

[5] K. Wada and T. Shibata, "Living with Seal Robots: Its Sociopsychological and Physiological Influences on the Elderly," IEEE Transactions on Robotics, vol. 23, no. 5, pp. 972–980, 2007.