Design of a Multimodal Psychological Companion Robot Based on Embedded AI

Luo Shunan[1,a,*], Sun Xiaoqing[1,b], Zhang Yucheng[1,c], Deng Yun[1,d]

[1]University of Science and Technology Liaoning, Anshan, China
[a]1026485105@qq.com, [b]1299095882@qq.com, [c]1066965502@qq.com, [d]3091576143@qq.com
**Corresponding author*

Abstract: *To address the mental health needs of groups such as left-behind children, long-term hospitalized patients, and the elderly living alone, as well as the limitations of traditional psychological services including uneven resource distribution and delayed response, an AI emotional interaction and psychological assessment companion robot based on an embedded platform was designed. The robot adopts a "local terminal+cloud/PC+mobile APP" collaborative architecture, with an embedded AI large model as its core, integrating multimodal perception, real-time psychological assessment, and autonomous mobility functions. At the hardware level, hierarchical design enables friendly interaction and flexible movement. At the software level, a "perception-assessment-feedback" closed-loop system is constructed, featuring core functions such as voice interaction, emotion recognition, dynamic psychological assessment, and visual report generation. Test results show that the robot achieves a speech recognition accuracy ≥96%, an emotion recognition accuracy ≥92%, and a system response time ≤0.8 seconds. It can provide continuous psychological support in various scenarios such as schools, medical institutions, and homes, effectively alleviating the psychological distress of target groups and offering an intelligent solution for mental health management.*

Keywords: *Embedded AI, Emotional Interaction, Psychological Assessment, Companion Robot, Multimodal Data*

1. Introduction

With social transformation and population mobility, the number of special groups such as children, hospitalized patients, and the elderly living alone continues to increase, and their mental issues are becoming increasingly prominent[1]. These groups commonly face problems like emotional deprivation, insufficient support, and delayed response to psychological needs, which can easily lead to emotional disorders such as loneliness, anxiety, and depression, or even psychological illnesses[2]. However, traditional psychological services rely on manpower, have limited coverage, slow response times, and high costs, making it difficult to provide continuous, lightweight, and timely support, especially in resource-scarce or closed environments. Therefore, researching and designing an AI emotional interaction and psychological assessment companion robot based on an embedded platform, capable of multi-scenario adaptation, real-time response, and continuous companionship, has become a market demand to fill the gap in mental health services for special groups and to build a human-machine collaborative psychological support system[3].

2. Hardware System Design

The hardware system of the psychological companion robot adopts a hierarchical collaborative architecture design, divided from top to bottom into the interaction layer, the core processing layer, and the functional execution layer. The interaction layer integrates a high-sensitivity microphone array and a friendly plush shell, responsible for collecting user voice and creating a pleasant interactive experience. The core processing layer centers around the high-performance STM32U5 main control chip and encrypted storage module, enabling multimodal data fusion reasoning and user historical data management. The functional execution layer utilizes DC geared motors, encoders, and infrared and ultrasonic sensors to drive the robot for precise movement and autonomous obstacle avoidance. These three layers collaborate efficiently through standardized interfaces, collectively supporting the robot's real-time perception, intelligent decision-making, and flexible companionship capabilities in complex scenarios. The overall system architecture is shown in Figure 1.

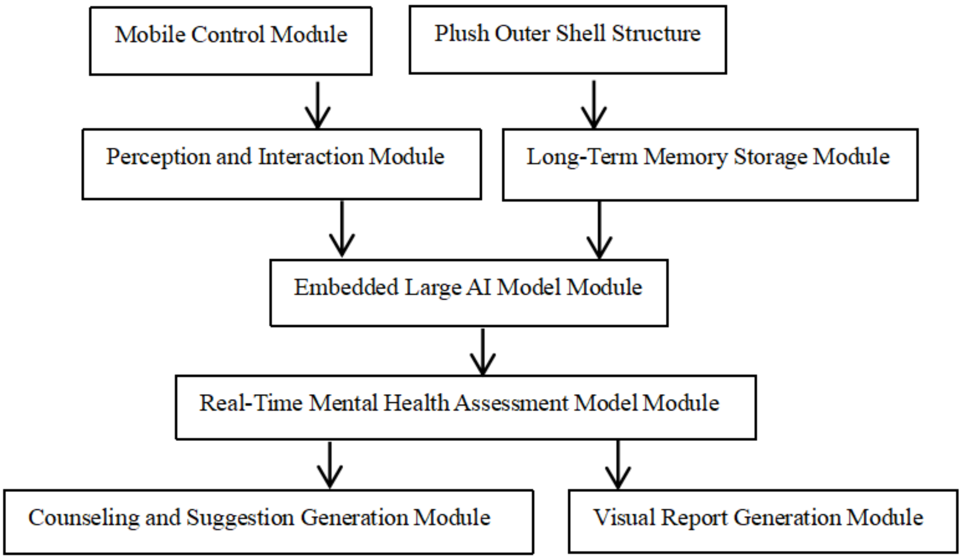
Mobile Control Module
Plush Outer Shell Structure
Perception and Interaction Module
Long-Term Memory Storage Module
Embedded Large AI Model Module
Real-Time Mental Health Assessment Model Module
Counseling and Suggestion Generation Module
Visual Report Generation Module

Figure 1: Overall Module Framework Diagram

2.1. Robot Hardware Main Body

To endow the robot with core capabilities of real-time emotional interaction, accurate psychological assessment, and active, flexible companionship, the design employs an integrated hardware architecture with a multi-layer stacked structure carrying the core modules. The core of the main body is the STM32U575ZITxQ main control chip based on the ARM Cortex-M33 core. Leveraging its powerful performance with a main frequency up to 160MHz, 2MB Flash storage, 786KB SRAM, and rich communication interfaces, it efficiently supports real-time fusion of multimodal data, inference of lightweight AI models, and multi-task collaborative scheduling.

The internally integrated intelligent power management module adopts a dual-power supply design. A high-capacity lithium polymer battery, combined with high-efficiency DC-DC step-down circuits and low-power LDO linear voltage regulators, provides stable and clean 3.3V and 5V voltages for the core processor, sensors, and actuators, respectively. Charging management chips enable fast charging and comprehensive protection, ensuring safe and reliable operation of the system under the endurance requirements of ≥12 hours standby and ≥6 hours continuous interaction.

2.2. Various Sensors

To achieve multidimensional perception of the environment and user state, this robot integrates four core types of sensors: a high-sensitivity microphone array for auditory perception and voice interaction, an infrared tracking sensor for path identification and following, an ultrasonic sensor for distance detection and obstacle avoidance, and an embedded language module responsible for speech signal processing and understanding.

The quad high-sensitivity microphone array integrated into the robot uses a Uniform Circular Array (UCA) configuration. Based on Beamforming and Generalized Cross-Correlation with Phase Transform (GCC-PHAT) sound source localization algorithms, it achieves spatial selective enhancement and noise suppression of target speech. This array features a high signal-to-noise ratio of ≥65dB and a wide frequency response range of 300Hz–8kHz. It can effectively improve speech intelligibility in complex acoustic environments and supports feature extraction such as Mel-frequency cepstral coefficients (MFCC) and prosodic features, providing robust acoustic input for backend semantic parsing and multidimensional emotion recognition models. It is the core sensing component for achieving precise human-robot emotional interaction.

The TCRT5000 infrared tracking sensor is a discrete environmental perception module based on reflective infrared detection. Its core consists of an infrared emitter and a photoelectric receiver, which emits infrared beams at a specific wavelength (typically 940 nm) and detects reflected signals from surfaces. It distinguishes between colors or materials—such as black-and-white boundaries—by their reflectivity differences, outputting corresponding digital switch signals.With an effective detection range of 1–8 cm, a response time ≤2 ms, and strong ambient-light immunity, the sensor serves as a discrete path-tracking unit in the navigation system. It reliably detects pre-laid ground guides—like black electrical tape or marked tracks—and provides binary feedback for precise local path following, enabling robust yet low-complexity autonomous tracking in structured environments. It is a key sensing component for lightweight positioning and guidance.The infrared line tracking module diagram is shown in Figure 2.

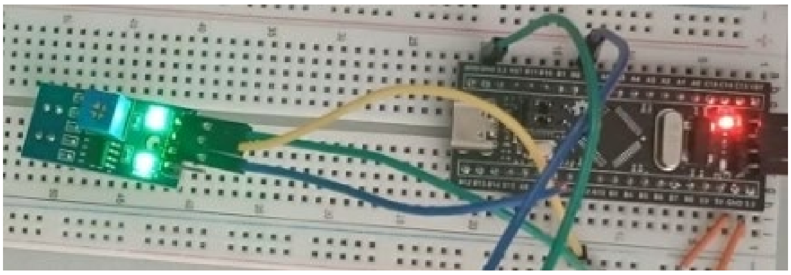

Figure 2: Infrared Line Tracking Module Diagram

The HC-SR04 ultrasonic ranging sensor employs the Time-of-Flight (ToF) ranging principle. It emits 40kHz ultrasonic pulses and precisely measures the echo time after encountering an obstacle, calculating the target distance based on the speed of sound propagation in air. This sensor has an effective ranging range of 2cm–400cm, a theoretical accuracy of up to ±3mm, and a horizontal detection angle of approximately 15°, enabling reliable detection of obstacles within a frontal sector area. In the system's real-time obstacle avoidance and navigation decision-making layer, this sensor continuously provides a high-frequency (typically >10Hz) distance data stream. When an obstacle distance below a preset dynamic safety threshold (e.g., adjustable between 20cm–50cm) is detected, the main control system, based on multi-sensor fusion results, immediately triggers obstacle avoidance behaviors such as local path replanning or emergency stop, thereby ensuring movement safety and autonomy in complex dynamic environments. It is the core "forward-looking" perception unit for the robot's real-time environmental perception and reaction. The Ultrasonic Ranging Module Diagram is shown in Figure 3.

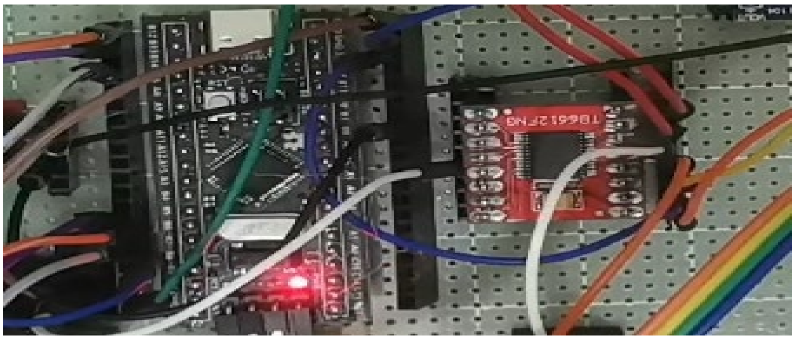

Figure 3: Ultrasonic Ranging Module Diagram

The embedded language module is built upon the Smart Gongyuan JX-A7T1 dedicated voice chip, integrating high-precision speech recognition, natural language understanding, and dialogue management functions. It supports a dual-mode operation mechanism combining offline wake-word interaction and online large language model collaboration. This module can not only process voice signals collected by the microphone array in real-time, enabling multi-turn contextual dialogue and emotionally responsive reply generation, but also synchronously extract acoustic features (such as tone, speech rate) from the voice as input for emotion analysis. Through UART/SPI interfaces, it efficiently collaborates with the main control system, thereby becoming the core processing hub for realizing the "perception-understanding-response" intelligent interaction closed loop. The Voice Module Diagram is shown in Figure 4.

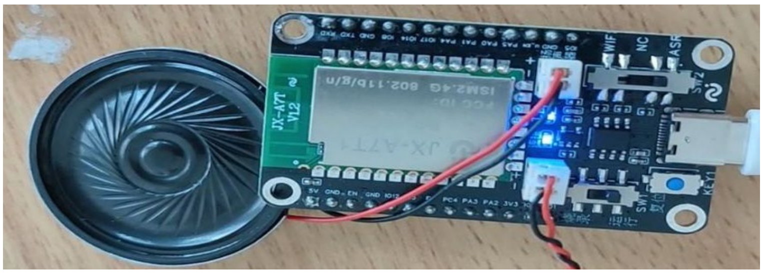

Figure 4:Voice Module Diagram

3. Software System Design

To enable the robot to achieve stable and reliable emotional interaction, real-time and accurate psychological assessment, and active, flexible personalized companionship, the software architecture of this system centers on an intelligent "perception-understanding-assessment-feedback" closed loop. It adopts a layered decoupling and modular design philosophy, constructing a cross-platform collaborative software system. Under the scheduling of an embedded real-time operating system (RTOS), the system ensures the efficient and reliable operation of each functional module and relies on a cloud collaboration platform for data storage, analysis, and visualization.

3.1. Main Module Design

The User Interaction Management Module serves as the core interface for direct interaction between the system and the user, undertaking the critical responsibilities of multimodal input/output scheduling and interaction state maintenance. Centered around a multi-turn dialogue manager, this module maintains complete dialogue context, coordinates the real-time workflow of the Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) engines, and manages non-verbal feedback channels such as expression lights and screen displays. By optimizing the signal processing pipeline and resource scheduling strategies, this module ensures low latency and high fluency throughout the entire chain from voice capture and semantic understanding to multimodal feedback.

The AI Agent and Psychological Assessment Dual-Core Module serves as the intelligent core of the system, comprising two tightly coupled sub-modules. The first is the AI Agent sub-module, which integrates a lightweight domain-adapted large model. It performs deep fusion of voice and text features to interpret user emotional intent, generates empathetic and personalized dialogue responses by leveraging long-term memory archives, and concurrently outputs structured emotional and semantic features to the assessment subsystem. The second is the Dynamic Psychological Assessment sub-module, which builds a multidimensional quantitative model grounded in cognitive-behavioral theory. By analyzing the incoming data stream in real time and applying weighted fusion along with time-series algorithms, it dynamically computes the user's psychological state index. When persistent anomalies or significant fluctuations are detected, the module automatically issues tiered warning signals, thereby supplying a real-time decision-making basis for precise counseling and external intervention.

The Task Execution and Data Coordination Module is the core hub connecting the system's decision-making layer with the physical execution end, undertaking the dual responsibilities of instruction parsing/execution and full-lifecycle data governance. Its task execution engine is responsible for receiving and parsing high-level instructions, precisely scheduling underlying hardware resources (such as motor drivers, audio modules, etc.) to complete tasks like autonomous navigation, media playback, and execution of specific behavior scripts. Simultaneously, the data coordination manager performs real-time encryption, cleaning, and structured storage of local interaction data, achieving reliable synchronization with the cloud platform through secure protocols like MQTT over TLS. The cloud provides data persistence, distributed analysis, and model iterative training support, offering continuous data services to front-end applications through standardized APIs.

The Visualization Application Service Module provides customized data interaction and management interfaces for users in different roles through a dual-end collaborative architecture comprising a mobile APP and a Web management backend. The mobile APP, as the core user terminal, integrates a lightweight AI dialogue interface and a visual dashboard. Utilizing core functions like ChartGeneration(), it converts multidimensional psychological assessment data into intuitive charts such as emotion curves and radar maps, significantly enhancing users' awareness of their own mental state, while also providing guardians with a remote monitoring and warning reception portal. The Web backend for institutional management offers functions including device cluster management, macro data dashboards, group psychological report generation, and warning summarization. It comprehensively supports the large-scale deployment and centralized operation/maintenance of robots, empowering professionals to conduct data-driven group psychological monitoring and intervention decision-making.

3.2. Key Algorithms

Embedded Lightweight AI Large Model Fusion Technology addresses the challenge of complex interaction under the computational constraints of embedded platforms. It employs techniques such as model pruning, knowledge distillation, and quantization to compress and optimize the base large model. Furthermore, a multimodal feature early fusion algorithm is designed: text features from speech recognition are fused with acoustic emotion features (e.g., MFCC, prosodic parameters) extracted directly from raw audio at the model's embedding layer. This enables the model to jointly understand linguistic content and paralinguistic information, thereby significantly improving emotion recognition accuracy and enhancing the depth of contextual dialogue comprehension[4].

To address the limitations of traditional static models, this paper proposes a time-series-based dynamic psychological assessment algorithm. It incorporates a temporal attention mechanism that analyzes not only current interactions but also links them with recent historical data to model continuous psychological changes[5].Using a sliding window approach, the algorithm computes short-term variations and long-term trends of each psychological dimension in real time. If any dimension score remains below a threshold or shows sharp decline, graded alerts are triggered—shifting assessment from static and discrete to dynamic and continuous monitoring. See Tables 1 and 2 for assessment criteria and result interpretations.

Table 1: Psychological Assessment Criteria

Emotional State	Emotional State For most of today, your mood was: 1=Very low/irritable; 2=Occasionally low/irritated; 3=Calm; 4=Occasionally happy; 5=Fairly happy
Sleep Quality	Sleep Quality Last night's sleep: 1=Insomnia (<4 hours)/restless with frequent awakenings; 2=Light sleep (4-6 hours); 3=Normal (6-8 hours); 4=Fairly deep sleep; 5=Very sound sleep
Energy & Motivation	Energy & Motivation Motivation to complete daily tasks (e.g., study, work, chores) today: 1=Completely unmotivated; 2=Barely managed to finish; 3=Completed normally; 4=Proactively completed; 5=Completed efficiently
Interpersonal Interaction	Interpersonal Interaction Feelings during interactions with others (family, friends, colleagues) today: 1=Strongly resistant/tense; 2=Managed with difficulty; 3=Calm; 4=Relaxed; 5=Pleasant
Stress Coping Ability	Stress Coping Ability Reaction to minor setbacks (e.g., disrupted plans, small mistakes) today: 1=Crumbled/self-blaming; 2=Irritated but held it in; 3=Handled calmly; 4=Adjusted quickly; 5=Resolved positively

Table 2: Interpretation of Psychological Assessment Results

Overall Score	Interpretation
5–10 points	Poor current psychological state, with possible significant distress (e.g., emotional breakdown, insomnia). Attention should be given to specific issues (e.g., interpersonal conflicts), or further discussion with me for details is recommended.
11–15 points	Psychological state is generally stable, but certain areas (e.g., sleep/motivation) may require adjustment.
16–15 points	Psychological state is good, with balanced emotions and social functioning.
Low score in a single dimension (≤2 points)	Indicates potential distress in that area (e.g., a consistent score ≤2 in "Sleep Quality" suggests the need to consider possible stressors or a tendency toward sleep disorders).

To ensure the end-to-end delay from voice input to behavioral feedback is strictly controlled within ≤0.8 seconds, the system implements multi-layered optimizations at the software level. First, a parallel audio-stream processing architecture pipelines and parallelizes stages such as voice capture, front-end noise reduction, and endpoint detection, significantly reducing signal preprocessing time. Second, a high-priority task scheduling strategy in the embedded real-time operating system (RTOS) grants the highest priority to threads responsible for perception capture and intelligent decision-making, preventing critical-path tasks from being blocked. Finally, a memory-resident mechanism for key data preloads and keeps frequently used dialogue templates, assessment model parameters, and other high-access data in memory, minimizing latency fluctuations caused by I/O operations. These measures work together to systematically ensure the real-time performance and determinism of interactive responses from three dimensions: processing flow, task scheduling, and data access.

3.3. Data Flow and Workflow

The system workflow begins with user-initiated interaction. The voice signal, after being captured by the microphone array and denoised, is sent to the language module for real-time recognition and text conversion. Both the text and the original audio features are simultaneously sent to the AI Agent for fusion understanding[6]. The AI Agent, on one hand, generates response content to drive speech synthesis and, on the other hand, outputs emotion labels and semantic summaries to the assessment module. The assessment module updates the psychological state score by combining the current data with the user's historical profile. Based on the scoring results, it decides whether to incorporate counseling discourse into the dialogue or send alerts to guardians via the APP[7]. All interaction data is encrypted; a portion is stored in the local long-term memory module, and another portion is asynchronously uploaded to the cloud for model iteration and remote visualization.

4. Conclusion

The AI emotion companion robot developed in this study is designed for scenarios with limited mental health resources or lasting companionship needs, such as rural schools, children's hospitals, and senior care homes[8]. It engages users in natural conversation, uses built-in AI for real-time emotion analysis, and moves autonomously to accompany daily activities[9]. Assessment reports are provided via a mobile app. This solution offers a low-cost, accessible, and sustainable form of psychological support. It helps address gaps in traditional services and shows strong potential for practical use in supporting mental well-being and improving public service efficiency[10].

References

[1] Nygaard M, Andersen S, Toftager M, et al. Screen use, physical activity, and sleep among adolescents: A latent class analysis of 4,421 Danish adolescents' movement patterns and the correlation with mental health[J].Mental Health and Physical Activity,2026,30100726-100726.

[2] Stewart C H, Talpur A A, Govera P, et al. Effect of Childhood Exposure to Domestic Violence and Abuse on Adult Relationships: A Mental Health Nurse Perspective.[J].Journal of psychosocial nursing and mental health services,2025,1-8.

[3] Giel E K ,Behrens C S ,Zipfel S .From mirror to mind: body dissatisfaction and mental health.[J].The lancet. Psychiatry,2026,13(1):5-6.

[4] Costantini I ,Eley C T ,Pingault B J , et al.Longitudinal associations between adolescent body dissatisfaction, eating disorder and depressive symptoms, and BMI: a UK twin cohort study.[J].The lancet. Psychiatry,2026,13(1):37-46.

[5] Verdugo L J ,Nobles T ,Herting R J , et al.Socioeconomic factors and mental health among young Asian adults in the United States.[J].Ethnicity & health,2025,1-21.

[6] Burke W C, Lanni R S, Firmin S E, et al. Resilience in Transitional Age Youth Experiencing Homelessness: The Role of Social Connectedness[J].JAACAP Open,2025,3(4):1016-1024.

[7] Frye S W, Ward S, Mauriello D, et al. Psychosocial profiles of autonomic dysfunction[J]. Autonomic Neuroscience: Basic and Clinical,2025,262103365-103365.

[8] Sanatkar S ,Lipscomb R ,Bower M , et al.Social Determinants of Recovery from Work-Related Psychological Injury After Sick Leave Absence: Examining Employee and Manager Perspectives[J]. Journal of Occupational Rehabilitation,2025,(prepublish):1-16.

[9] Zhou Z, Yang Z, Song Z, et al. Design and development of a novel desktop robot for orthodontic archwire forming[J].Robotics and Autonomous Systems,2026,196105243-105243.

[10] Lee W, Ryu S, Park G, et al. Mobile robot design with linkage-based reaction wheel mechanism for horizontal–vertical force transmission[J].Intelligent Service Robotics,2025,18(6):1-13.