

Multimodal Artificial Intelligence for Elderly Care Robots in Perception Interaction Decision Making and System Deployment

Lizhi Wang

*Suzhou High School of Jiangsu Province, Suzhou, China
leonardo.wang@szzx-intl.cn*

Abstract. With the rapid growth of the aging population in the world, elderly care systems are facing increasing pressure and challenges in healthcare services, daily assistance, and emotional support. Artificial intelligence (AI) has been widely integrated into elderly care robots to enhance autonomy, perception, interaction, and decision-making capabilities. This paper reviews recent advances in AI for elderly care robots, focusing on perception, natural language processing, machine learning, affective computing, and decision-making. In addition, it examines key studies, compares primary technical approaches, and identifies key challenges, including privacy risks, ethical issues, system reliability, and human-robot trust. Based on recent studies, this study explains how AI improves elderly care quality and identifies gaps in real-world deployment and long-term performance. The results reveal that AI improves personalization and autonomy, while major challenges remain in safety validation, ethical governance, and multimodal interaction performance, which need to be addressed to enable scalable and sustainable elderly care solutions.

Keywords: Elderly Care Robot, Artificial Intelligence, Human-Robot Interaction, Healthcare Robotics, Ethical governance

1. Introduction

With the accelerated development of global aging population, the demand for elderly care is continuously increasing, while the shortage of nursing personnel and the pressure on medical resources are constantly intensifying, which has driven the rapid development of intelligent elderly care technologies. Current research indicates that elderly care robots integrating artificial intelligence and robotics technology have become an important development direction in this field. Relevant studies have made progress in multiple perception modalities including computer vision, speech recognition, and physiological signal analysis, as well as autonomous decision-making and human-computer interaction, such as visual recognition based on CNN and Transformer, large language model interaction based on NLP, and real-time health monitoring systems based on edge computing [1, 2]. However, existing studies still have limitations, like weak robustness in complex unstructured home environments, poor co-optimization between algorithms and embedded hardware, incomplete multimodal fusion methods, and the absence of a unified framework for data

privacy and ethics. Based on this, this paper centers on an AI-powered elderly care robot and explores the relationships among its core AI technologies, system architecture, functional mechanisms, and performance optimization strategies. By reviewing the existing studies, it summarizes and compares them to identify main technical paths and research gaps, and proposes future directions and optimization strategies from engineering and application perspectives.

2. Technical fundamentals and system architecture

2.1. AI algorithms and perception systems

From a system architecture perspective, modern elderly care robots are typically built on a multimodal perception framework integrating computer vision (CV), natural language processing (NLP), and multi-source sensor fusion systems. For visual perception, convolutional neural networks (CNNs) are used for low-level feature extraction and local spatial representation, while vision transformers (ViTs) have become the dominant approach for complex scene understanding due to their global attention mechanism.

In the speech and language interaction module, the pre-trained models based on Transformer form the current mainstream technical path, which are used for unified modeling of speech command semantic parsing and unstructured expression; Recurrent Neural Networks (RNNs) and their variants are mainly applied to lightweight temporal modeling tasks, and their application scope is gradually shrinking. In the multimodal perception layer, heterogeneous sensors jointly model environment and human state. Visual data come from RGB-D and infrared cameras, environmental sensing from LiDAR and ultrasonic arrays, and physiological signals from PPG and wearables, forming time-series data such as heart rate, SpO₂, and body temperature. These signals require synchronization and feature alignment, then are fused for behavior recognition and health analysis [3]. At the system implementation level, multimodal perception is usually supported by the Internet of Things and embedded system architectures. Low-power sensing networks and edge computing nodes are used to enable data collection and real-time processing, meeting the requirements of long-term monitoring and real-time response [4]. In real elderly care scenarios, the design of service robot systems emphasizes the collaborative integration of perception, decision-making, and execution to ensure reliable operation in complex environments [5].

2.2. Embedded systems and integration architecture

The embedded system and integrated architecture deploy perception and decision-making algorithms on resource-limited edge hardware while enabling real-time coordination across modules, with ARM Cortex-A processors handling operating system execution, task scheduling, data synchronization, and system communication control, and an NPU-GPU heterogeneous unit accelerating local deep learning inference to meet real-time requirements [4]; at the hardware level, a hierarchical bus structure connects computing, sensing, and actuation units, while the main controller communicates with motor drives, mobile chassis, and peripheral modules via I2C, SPI, UART, and CAN to ensure deterministic latency and stable data transmission [5].

At the software integration level, the system is typically built on ROS/ROS2, forming a distributed runtime framework in which each functional module operates as an independent node. These nodes exchange information through publish/subscribe communication and service calls, enabling a loosely coupled architecture that facilitates modular expansion, fault isolation, and stable runtime behavior [5]. In practical deployment, this design also allows dynamic reconfiguration of

modules without disrupting overall system operation, which is particularly important for long-term assistive care tasks. Besides, for long-duration embedded medical scenarios, edge computing is often combined with low-power sensor networks to balance real-time responsiveness and energy consumption [3]. By offloading computation to nearby edge nodes while keeping sensing units lightweight and continuously active, the system can maintain stable monitoring and timely response under constrained power budgets.

3. Key functionalities and technical approaches

3.1. Health monitoring and remote data management

Health monitoring and remote data management form the core medical functions of elderly care robots, thus enabling continuous physiological data collection, stable transmission, and edge-cloud coordination. For long-term operation, the system adopts a layered data processing and transmission architecture, improving robustness under unstable networks while preserving real-time responsiveness.

In the data acquisition and local processing layer, the robot obtains multi-source physiological data via wearable devices and built-in sensing modules, and completes time alignment and data preprocessing in the local computing unit to ensure the continuity and consistency of the data flow. The preprocessed data is then transmitted to the robot's edge processing unit through low-power wireless communication methods (such as BLE or local Wi-Fi) [3]. In the data transmission and cloud interaction layer, the system adopts lightweight IoT communication protocols (such as MQTT) or secure HTTPS channels to encrypt and upload structured health data to the remote medical cloud platform for long-term storage and remote access [4]. This architecture embodies a typical edge-cloud collaborative processing model, where time-sensitive tasks are executed at the edge end, while long-term trend analysis and historical data modeling can be completed in the cloud. In terms of anomaly detection and alarm mechanism, the system combines real-time detection at the edge and analysis results from the cloud to assess the health status of continuous time series data. Upon detecting abnormal patterns, the system generates a local alert and simultaneously dispatches high-priority notifications to remote caregivers and family members, thereby establishing a multi-channel coordinated response mechanism [6].

3.2. Mobility and daily living support systems

The mobile and daily-life assistance system forms the basic functional layer of the elderly care robot and directly influences its autonomous operation in complex home environments [2]. In the motion control architecture, the chassis typically uses a hierarchical control strategy. Model predictive control (MPC) is used at the upper level for trajectory planning and predictive path optimization, where reference velocity or trajectory sequences are generated by solving constrained optimization problems over a finite horizon. In contrast, the lower layer relies on high-frequency PID control loops to perform real-time closed-loop regulation of motor speed and posture. And this forms a two-layer framework that decouples motion planning from low-level control.

For robotic arm control, multi-degree-of-freedom motion is typically realized by inverse kinematics (IK), which maps task space to joint space, and by trajectory interpolation methods that generate smooth continuous motion paths, supporting general grasping and manipulation tasks. In terms of autonomous navigation, the system builds an occupancy grid map of the environment via Simultaneous Localization and Mapping (SLAM), and then performs path planning tasks based on

this map. Generally, global path planning uses the A* algorithm to generate the optimal path, while local obstacle avoidance employs the Dynamic Window Approach (DWA) to adjust in real time to adapt to the changes in the dynamic home environment [5]. In terms of safety and emergency response, the fall detection is achieved through the fusion of visual skeleton estimation and inertial sensor data. When a fall event is detected, the system interrupts the current task and triggers path re-planning, enabling the robot to quickly reach the target location. Meanwhile, the remote communication module is activated to establish a connection with the caregivers, enabling remote assessment and intervention support [6].

3.3. Human-robot interaction and affective computing

The human-machine interaction (HRI) and emotion computing module in intelligent care robots aim to achieve the understanding of human intentions and the generation of interaction strategies under multi-modal input conditions. This system constructs a unified interaction representation by integrating voice, vision, and physiological signals, thus supporting semantic understanding and behavior decision-making in complex environments [5].

In the specific implementation, voice interaction acts as the main input channel. Firstly, a microphone array collects multi-channel signals, and beamforming technology is employed to enhance the signal-to-noise ratio of the target sound source. On this basis, the system converts the voice signal into a structured semantic representation via the automatic speech recognition (ASR) and natural language understanding (NLU) modules, and inputs it into the task parsing and decision-making module to achieve the mapping from instructions to actions. Based on voice and semantic information, the system further integrates a multimodal emotion perception mechanism to improve the representation of the interaction state. The visual modality encodes through facial action units (FAU) and expression features, the speech modality represents emotion-related characteristics based on prosodic features, and the physiological modality reflects the user's state using time series indicators like heart rate variability (HRV). These heterogeneous information, after being synchronized in time, enter the multimodal fusion module to form a unified state estimation result. Based on the above semantic and emotional joint representation, the system enters the decision-making and behavior generation stage. For long-term operation, the system uses a layered data pipeline, improving robustness under unstable networks while maintaining real-time performance [6, 7].

4. System performance and optimization strategies

4.1. Performance metrics and evaluation

In order to conduct a quantitative assessment of the engineering feasibility and operational efficiency of elderly care robots, a standardized performance index system needs to be adopted, and a comprehensive analysis should be carried out in combination with a unified testing method. Generally, the evaluation indicators of such systems can be summarized into four dimensions: algorithm accuracy, response delay, energy consumption efficiency, and system reliability [5].

In terms of algorithm accuracy, the performance of the robot in tasks such as target recognition, behavior understanding, and detection of falls or health abnormalities is mainly evaluated. Common metrics include the mean average precision (mAP) for target detection, the F1-score for behavior recognition, and the sensitivity and specificity for safety-critical detection tasks. Among these, specificity is particularly important in practical applications, as it helps to reduce false alarms and

avoid the "alert fatigue" problem in the nursing system [6]. In terms of response latency, the evaluation system measures the end-to-end time cost from the sensor input to the decision output and the execution action. For tasks involving physical interaction and security control, this delay typically needs to be controlled within the millisecond range to ensure the real-time responsiveness in dynamic environments [6]. From an energy efficiency perspective, system power consumption is quantitatively evaluated across different operating workloads, including standby, mobile navigation, and real-time inference. Besides, current and thermal behavior are analyzed, and battery life per charge is measured to assess overall energy efficiency [4]. In terms of system reliability, the average fault interval time (MTBF) and data communication packet loss rate are usually adopted as the core evaluation indicators to measure the stability and fault-tolerant capability of the system during long-term operation and in unstructured environments [4].

4.2. Deployment-oriented algorithm-hardware co-optimization

In practical deployment scenarios, performance optimization of elderly care robots relies on improvements in individual algorithms and also on a coordinated optimization mechanism spanning algorithm design, hardware resource allocation, and runtime scheduling. Under this framework, the algorithm optimization reduces computational cost using model compression, pruning, quantization, knowledge distillation, and early-exit inference, enabling adaptation to resource-limited edge environments. However, the selection of these methods is usually directly related to the task risk level: safety-critical tasks such as fall detection, obstacle avoidance, and emergency response must rely on low-latency and high-reliability local inference, while with lower real-time requirements can be partially offloaded to the edge or cloud when network conditions are stable, achieving a dynamic balance of computing resources [4].

At the level of hardware and runtime collaboration, this system further manifests as the task mapping and scheduling optimization of heterogeneous computing resources. The underlying motor control and sensor polling are typically handled by the microcontroller to ensure the real-time and determinism of the control loop; the main CPU is responsible for system-level scheduling and cross-module coordination tasks; while the computationally intensive models are accelerated on GPUs, NPU or DSP based on the hardware conditions. Through the DMA data transmission mechanism, hierarchical buffering structure and real-time scheduling strategy, the system can avoid high-frequency data streams from blocking the safety control link [5]. In practical deployment, these optimization strategies must also address real-world constraints, including wireless network instability, lighting variations, sensor occlusion, battery aging, and thermal limitations. Thus, the algorithm and hardware collaborative optimization must be evaluated through hardware-in-the-loop (HIL) testing and real-world validation, and the end-to-end latency, peak memory usage, single inference energy consumption, thermal stability, and fault recovery capabilities are used as the comprehensive evaluation criteria [3].

4.3. Technical outcomes and limitations

From a system optimization perspective, hardware-software co-design is widely employed to maintain stability and real-time performance in long-term operation. In prototype systems, tasks such as vital sign monitoring and fall detection achieve both low latency and high accuracy. Through edge computing optimization and lightweight models, energy consumption is reduced, improving continuous operation in home and clinical settings and supporting basic engineering validation [1, 8].

Nevertheless, during the process of transitioning from the prototype system to large-scale deployment and commercial application, there are still several key technical bottlenecks. In terms of computing and thermal management, as the demand for multi-modal basic models and large-scale pre-trained models on the edge increases, the contradiction between computing resources and heat dissipation capacity of the system becomes increasingly prominent. Especially when simultaneously processing high-frame-rate visual data and language model inference, this problem becomes more significant. The robustness in the unstructured environment remains a key factor limiting the reliability of the system. Elderly living spaces typically exhibit high dynamic complexity, such as changes in lighting conditions, unpredictable moving obstacles, and differences in ground materials. These factors can degrade visual perception performance and further destabilize SLAM systems, leading to issues such as accumulated localization error and map drift. From a data and ethics perspective, elderly care robots collect sensitive health and environmental data, which raises concrete security and compliance requirements. Although existing studies on home assistive technologies exist, long-term deployment in real care settings is still limited by ethical approval and user acceptance [8]. In addition, robots may increase user dependence and raise concerns about replacing human caregivers [9]. Meanwhile, during sensing, monitoring, and cloud transmission, privacy and user autonomy are directly affected. Furthermore, a unified ethical governance framework is still missing, which limits reliable deployment [10].

5. Conclusion

This review shows that AI-based elderly care robots have reached a relatively mature technical stage. Computer vision, natural language processing, and machine learning respectively underpin key functions such as fall detection, social interaction, and health monitoring. However, several challenges remain. From a computational standpoint, deploying complex deep learning models on resource-constrained embedded hardware still faces significant engineering bottlenecks. From a social perspective, continuous collection of sensitive health and environmental data raises serious privacy concerns, while the lack of standardized security and verification protocols further weakens user trust in these systems. Looking ahead, future development should focus on three closely linked directions: designing lightweight, edge-deployable AI architectures to meet real-time requirements without heavy reliance on cloud computing; developing robust multimodal interaction systems that fuse physiological and behavioral signals for more accurate emotion understanding; and strengthening interdisciplinary collaboration to build global ethical frameworks and safety standards. Only through the coordinated advancement of these areas, rather than isolated efforts, can current laboratory achievements be translated into scalable and reliable solutions for real-world elderly care applications.

References

- [1] Wong, K. L. Y., Hung, L., Wong, J., et al. (2024). Adoption of artificial intelligence-enabled robots in long-term care homes by health care providers: Scoping review. *JMIR Aging*, 7, e55257. <https://doi.org/10.2196/55257>
- [2] Padhan, S., et al. (2023). Artificial intelligence (AI) and robotics in elderly healthcare: Enabling independence and quality of life. *Cureus*, 15(8). <https://doi.org/10.7759/cureus.42905>
- [3] Rathnayake, D., Arawa, P., & Ekanayake, T. (2025). AI in wearable embedded systems for healthcare monitoring: A review. *Sri Lankan Journal of Applied Sciences*, 4(1), 41-54. <https://doi.org/10.4038/sljoas.v4i1.15>
- [4] Shirvanian, N., Shams, M., Rahmani, A. M., et al. (2025). The Internet of Things in elderly healthcare applications: A systematic review and future directions. *IEEE Access*, 13, 71335-71373. <https://doi.org/10.1109/ACCESS.2025.3562147>

- [5] Riek, L. D. (2017). Healthcare robotics. *Communications of the ACM*, 60(11), 68-78. <https://doi.org/10.1145/3127874>
- [6] Cesário, L. D. C., Barbosa, P., Miguel, P. A. C., et al. (2025). Service robots in caring for older adults: Uncovering the current conceptual and intellectual structures and future research agenda. *Archives of Gerontology and Geriatrics*, 131, 105755. <https://doi.org/10.1016/j.archger.2025.105755>
- [7] Shen, J., Yu, J., Zhang, H., et al. (2025). Artificial intelligence-powered social robots for promoting physical activity in older adults: A systematic review. *Journal of Sport and Health Science*, 14, 101045.
- [8] Daly, J., Rondón-Sulbarán, J., Quinn, E., et al. (2017). A systematic review of electronic assistive technology within supporting living environments for people with dementia. *Dementia*, 18(7-8), 2371-2435.
- [9] Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141-161. <https://doi.org/10.1007/s11023-006-9030-6>
- [10] Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*, 14(1), 27-40. <https://doi.org/10.1007/s10676-010-9234-6>