Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3096

AFFECTIVE COMPUTING AND MULTIMODAL INTERACTION FOR SOCIAL HUMANOID ROBOTS

ShuoPei Yang, NaNa Wang*
Lingjing Jushen (Ningbo) Electronic Technology Co., Ltd., Ningbo 31500, Zhejiang, China.
Corresponding Author: NaNa Wang, Email: 2373770@qq.com

Abstract: This study focuses on enhancing affective computing and multimodal interaction capabilities in social humanoid robots to improve natural communication experiences between robots and humans. By constructing a system framework that integrates affective computing with multimodal interaction, it addresses the key challenges of insufficient accuracy in emotion recognition and lack of naturalness in emotional expression. In the research design, we established a comprehensive system architecture encompassing multimodal emotional feature extraction, emotion state inference algorithms, and emotional expression strategies. At the interaction implementation level, a multimodal interaction system comprising perception, decision-making, and execution layers was designed to ensure effective processing of multi-source information such as voice, vision, and touch. Experimental results demonstrate significant improvements in key metrics including emotion recognition accuracy, interaction fluency, and user experience. Statistical analysis further validates the effectiveness of the proposed method. This research not only provides innovative technical solutions for social robots but also contributes substantially to both theoretical development and practical applications in the field of human-computer interaction.
Keywords: Humanoid robot; Emotion recognition; Human-computer interaction; Multimodal fusion

1 INTRODUCTION

The development of social humanoid robots, a significant branch of artificial intelligence (AI), has witnessed remarkable progress in recent years. With advancing technological maturity, these robots are increasingly being deployed across various sectors such as services, education, and entertainment. Statistical projections indicate that the global market for social humanoid robots is expected to grow at a double-digit annual rate in the coming years, underscoring their vast application potential. Affective computing, a key technology for enhancing the interactive capabilities of social humanoid robots, aims to endow them with the ability to understand and express emotions. This field encompasses not only the recognition of human emotional states but also the study of emotion generation and expression mechanisms. Research in affective computing is crucial for elevating the intelligence level of robots and achieving genuine human-robot emotional communication. Concurrently, the demand for multimodal interaction is becoming increasingly prominent in human-computer interaction research. Traditional unimodal interaction methods often fall short in complex scenarios. The integration of visual, auditory, tactile, and other modalities can provide a richer and more natural interactive experience. Studies suggest that multimodal interaction significantly improves the accuracy and efficiency of information transfer, making it a vital research direction for social humanoid robots. However, significant challenges remain in affective computing and multimodal interaction for these robots. These include the need for improved accuracy in affective model construction and emotion recognition techniques, the ongoing development of multimodal data processing and fusion technologies—particularly regarding the effective integration of information from different modalities—and the need for further optimization of emotional expression strategies to enable more natural and realistic emotional communication.

This research holds substantial theoretical value and practical significance. Theoretically, it deepens the understanding of affective computing mechanisms and multimodal interaction, offering new perspectives and technical pathways for the AI field. The construction of affective computing models allows for the exploration of the complex relationship between human emotions and machine behavior, thereby providing robots with more nuanced emotional expression and interaction capabilities. Research on multimodal interaction technology helps overcome the limitations of traditional unimodal interfaces, facilitating the fusion of various perceptual information to enhance both the robot's environmental perception and the naturalness and effectiveness of interaction. In terms of application prospects, this study has the potential to accelerate the commercialization of robotic technology, offering intelligent solutions for various industries. For instance, in education, social humanoid robots can serve as auxiliary teaching tools; in healthcare, they can provide companionship and psychological support; and in domestic settings, they can assist the elderly. Specifically, by leveraging affective computing and multimodal interaction, these robots can better understand and meet user needs, adjusting their behavior and language based on recognized user emotional states to enable more natural and appropriate interaction, which is essential for enhancing their social competence and fostering harmonious human-robot coexistence.

To address the key scientific challenges, this study aims to construct an effective affective computing model for accurate emotion recognition and expression in social humanoid robots, and to optimize multimodal interaction technology to enhance interactive fluency and user experience. The corresponding hypotheses are that the effectiveness of the affective computing model directly influences interaction quality, and that optimized multimodal interaction will

significantly improve fluency and user experience. The paper is structured into seven main parts: introduction, literature review, research design, experiments and data analysis, discussion of results, conclusion, and references, providing a clear framework to explore these issues systematically.

2 THEORETICAL AND TECHNICAL FOUNDATIONS OF AFFECTIVE COMPUTING AND MULTIMODAL INTERACTION

2.1 Foundations of Affective Computing

Emotion is an important component of cognitive processes and has a profound influence on human behavior and decision-making. Affective computing, as a branch of artificial intelligence, aims to enable machines to understand, simulate, and respond to human emotional states. Mechanisms for emotion generation and expression are among the core issues in affective computing, involving how to construct models that produce appropriate emotional responses and how to convey those responses effectively. Mechanisms for emotion generation are typically grounded in affective models and psychological theories. An affective model is an abstract representation of human emotional states and can be rule-based or instance-learned. In rule-based models, emotional states are controlled by a set of rules and parameters that simulate the processes of emotion emergence and change[1]. For example, emotion-regulation theories can be translated into rules that guide the generation of emotions.In instance-learned (data-driven) models, emotion generation is achieved by learning from large amounts of labeled data. These models can identify associations between emotional states and specific contexts and generate corresponding emotional responses accordingly. Deep learning techniques—especially recurrent neural networks and long short-term memory networks—have shown strong capabilities in this area, enabling the handling of complex emotional states and contextual information.Emotion expression mechanisms focus on transforming generated emotional states into perceivable signals such as facial expressions, vocal prosody, and body gestures. In social humanoid robots, these signals are crucial for establishing emotional rapport with human users. For example, by mimicking human facial expressions, a robot can convey emotions such as empathy, happiness, or sadness.The visual modality plays a prominent role in emotional expression: robots use facial expressions and body movements to express emotions. Facial expression synthesis techniques employ computer graphics methods to produce facial animations corresponding to emotional states. Speech synthesis technologies adjust pitch, volume, and rhythm to convey different emotions. In addition, the integration of haptic and other modalities offers new avenues for emotional expression—for instance, by simulating tactile feedback on human-like skin, robots can provide richer affective interaction experiences.

Research indicates that affective computing models have steadily improved in accuracy for both emotion recognition and generation. For example, deep-learning–based emotion recognition systems have reached near-human accuracy in identifying facial expressions and vocal emotions. Nonetheless, affective computing still faces many challenges, including the diversity and complexity of emotional expression and adaptability across different cultures and contexts.In multimodal interaction, affective computing models must integrate information from multiple modalities to achieve more accurate emotion recognition and expression. This requires models capable of handling time-series data and establishing effective mappings across modalities. Additionally, affective computing models need to consider real-time performance to meet the fluency requirements of interactive scenarios.In summary, mechanisms for emotion generation and expression play a key role in affective computing. Although progress has been made, further research and innovation are needed to achieve more natural and effective affective interactions. Future work may explore more advanced affective models and new approaches that combine cognitive science with AI techniques to advance the field.

2.2 Multimodal Interaction Technologies

Haptic perception, as one of the important ways humans sense the external world, has significant value when integrated into multimodal interaction technologies. Studies show that haptic feedback can enhance user immersion and interaction experience, especially in applications such as virtual reality and augmented reality. The integration of haptics with other modalities involves technical challenges at multiple levels, including haptic signal acquisition, processing, rendering, and fusion with information from other modalities. First, in the processing of the haptic modality, researchers have developed various haptic sensors capable of accurately detecting and encoding tactile signals. For example, haptic sensors based on capacitive or resistive principles can convert tactile stimuli into electrical signals, which are then processed and rendered by specific algorithms. Moreover, haptic display technologies—such as haptic feedback gloves and haptic projection systems—are continually evolving and can provide users with intuitive tactile experiences.The fusion of auditory and haptic modalities is another key aspect of multimodal interaction. In fields like voice interaction and music production, combining haptic feedback with auditory interaction technologies can significantly improve user experience. For instance, haptic vibration feedback can enhance the accuracy of speech recognition, or provide tactile cues when using digital musical instruments to emulate the feel of traditional instruments.In the integration of visual and haptic modalities, research focuses on converting visual information into haptic signals. For example, haptic feedback devices embedded in virtual reality headsets enable users to feel tactile effects corresponding to virtual scenes, thereby increasing immersion. Additionally, combining computer vision techniques that recognize users' actions and expressions with haptic feedback allows systems to respond to users' emotional states in real time.One of the cores of multimodal interaction technology is information fusion and decision-making across modalities. This requires systems not only to process and interpret information from different modalities, but also to make intelligent decisions based on

context and user intent. For example, in social humanoid robots, integrating visual, auditory, and haptic information enables the robot to understand a user's emotional state and needs more accurately, thereby producing more natural responses.However, the technical challenges of integrating haptics with other modalities should not be underestimated. Precise acquisition and efficient processing of haptic signals demand complex algorithms, and individual differences in haptic perception among users increase implementation difficulty. In addition, the integration and ergonomic comfort of haptic devices remain important topics in current research[2].

Statistics show that multimodal interaction technologies have broad application potential in user experience, affective computing, and human–computer interaction. Despite existing technical bottlenecks and challenges, continued research and technological advances will bring new opportunities for the integration of haptics with other modalities and the development of multimodal interaction technologies.

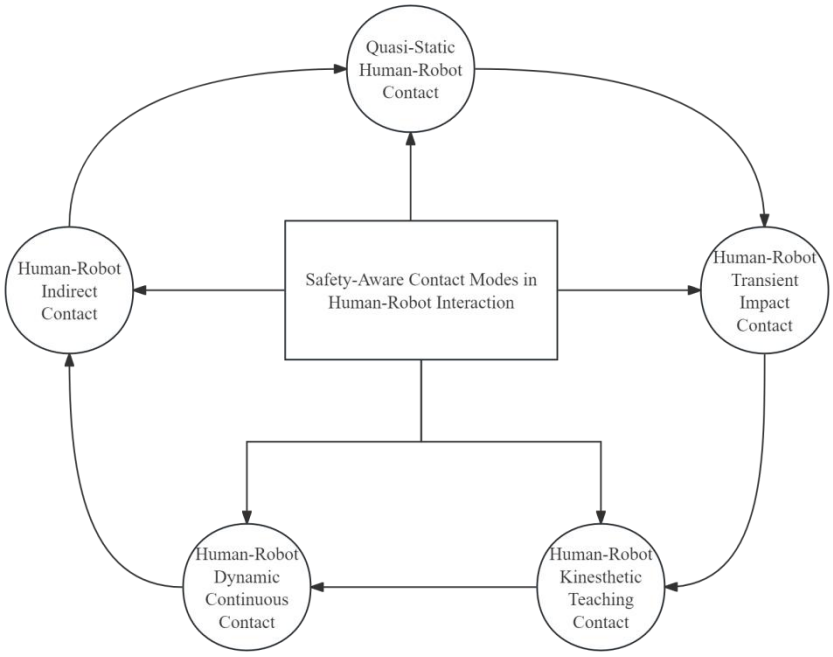

Figure 1 Safety Contact Mode for HRI

In the analysis of safe contact modalities for human–robot interaction, the focus is on the dynamic interaction process when a robot contacts the human body. This process encompasses not only the transmission and distribution of forces but also the dynamic evolution of contact points, the time-varying characteristics of contact forces, and the frictional properties of the contact interface. A thorough analysis of these factors enables a better understanding of safety risks during human–robot interaction and provides theoretical support for designing robots that are safer and more comfortable. The national standard GB/T 36008−2018 describes two contact modalities between humans and robots: quasi-static contact and transient contact. Quasi-static contact refers to contact between an operator and robot system components in which the operator's body parts may be trapped between a robot system moving part and another fixed or moving part of a robot unit. Transient contact refers to brief contact between an operator and robot system components in which the operator's body parts are not trapped by moving parts of the robot system and can rebound or withdraw from the contact. Based on analysis of human–robot contact modes, three additional modalities are identified: human–robot drag-teaching contact, human–robot dynamic continuous contact, and human–robot indirect contact. The safety contact modalities in human–robot interaction are illustrated in Figure 1. Human–robot drag-teaching contact typically occurs during teach-by-guiding programming, where the operator guides the robot's movement to record motion trajectories; this modality requires the robot to have a highly sensitive force-feedback mechanism to ensure precise execution of prescribed motions under operator guidance while preventing injury to the operator. Human–robot dynamic continuous contact involves sustained contact between the robot and the operator during task execution—such as in collaborative handling or assembly operations—where the robot must move synchronously with the operator, requiring strong dynamic response capabilities and robust safety control strategies[3].

2.3 Progress in Research on Social Humanoid Robots

Social humanoid robots have made significant research progress in recent years, but they have also encountered many bottlenecks and challenges during technological development. First, in terms of technical bottlenecks, robots' affective computing and multimodal interaction capabilities remain immature. Studies show that affective computing has limited accuracy when dealing with complex emotional states, particularly struggling with micro-expression recognition and fine distinctions among emotions. In addition, mechanisms for resolving redundancy and conflicts during multimodal

information fusion still need improvement.Second, hardware limitations of robots are another major challenge. Current humanoid robot hardware platforms lag far behind humans in strength, dexterity, and durability, which constrains their applicability in real-world environments. At the same time, issues of hardware integration and energy consumption affect robots' practicality and portability.Regarding affective models and representation methods, although various affective models have been proposed, constructing a model that accurately captures the complexity of human emotions remains difficult. Existing models are often based on simplified emotional theories and struggle to encompass the diversity and dynamics of human affect.In the research progress of multimodal interaction technologies, substantial advances have been made in processing the visual modality, especially in face recognition and expression analysis. However, auditory modality processing still faces challenges, such as degraded speech recognition accuracy in noisy environments and limitations in natural language understanding.The fusion of haptics with other modalities also faces technical challenges. Haptic feedback has important value in robot interaction, but current research is still at an early stage; how to convey information effectively while preserving tactile realism is an active research topic.Concerning the domestic research landscape, work on social humanoid robots in China has been progressively deepening, and several research teams have achieved notable results in affective computing and multimodal interaction. For example, some teams have successfully developed robot prototypes with emotion-recognition capabilities and carried out preliminary applications in specific scenarios.Nevertheless, technical bottlenecks and challenges persist. For instance, the naturalness and realism of robot emotional expression still need improvement, and the user experience during interaction requires further optimization. Moreover, research on the ethical and societal impacts of social humanoid robots is still insufficient; ensuring that robots' behavior conforms to social norms and ethical requirements and avoiding potential negative consequences will be essential in future work.Overall, research on social humanoid robots is advancing rapidly but still requires deeper investigation in technology, ethics, and social impact to promote sustainable development in this field.

2.4 Research Gaps and Innovations

In the fields of affective computing and multimodal interaction, despite substantial progress at home and abroad, many research gaps remain. First, existing affective models tend to focus on single-modality emotion recognition, and research on multimodal emotional feature fusion is still insufficient. For example, how to effectively integrate heterogeneous data from visual, auditory, and haptic sources to build a comprehensive and accurate multimodal affective state model is an important open problem. In addition, research on emotion expression strategies for social humanoid robots during actual interactions is relatively lacking; how to make robots better emulate human emotional expression to enhance the naturalness and authenticity of interaction is an urgent issue to address.The innovations of this study are several new ideas and methods proposed on top of existing research. First, this study proposes a deep-learning–based method for multimodal affective feature extraction that can effectively fuse affective information from different modalities to improve recognition accuracy and robustness. Second, it develops an affective state inference algorithm that combines cognitive psychology theory, enabling more accurate inference of users' emotional states and adaptive adjustment of the robot's interactive behavior. Additionally, this study designs a set of affective expression strategies tailored to social humanoid robots, which, by simulating human emotional expression habits, enhance the robot's affective interaction capabilities.At the frontier of international research on social humanoid robots, emphasis is mainly on technological innovation, while domestic research tends to focus more on applied exploration. Nevertheless, both international and domestic work face common technical challenges—for example, improving robots' perception in complex environments, optimizing emotion generation and expression mechanisms, and handling ethical and societal issues between robots and humans remain difficult problems.While addressing the above research gaps, this study also proposes the following innovations: first, the integration of multimodal affective feature extraction with affective state inference for the first time to enhance social humanoid robots' affective interaction capabilities; second, an affective expression strategy grounded in cognitive psychology to make robot interactions more natural and authentic; third, experimental validation demonstrating the feasibility of the theoretical models and methods proposed here, thereby offering new research directions for the further development of social humanoid robots[4].

3 DESIGN AND IMPLEMENTATION OF A SOCIAL HUMANOID ROBOT SYSTEM

3.1 Overall Research Framework

The overall research framework designed in this study aims to realize a comprehensive social humanoid robot system that achieves natural communication with human users via multimodal interaction technologies. The framework's core is a modular design that ensures effective collaboration among subsystems and standardized interface design, thereby improving system scalability and maintainability.In the system architecture overview, we adopt a layered design dividing the system into perception, decision, and execution layers. The perception layer is responsible for collecting and processing input from different modalities, such as visual, auditory, and haptic data. The decision layer uses information provided by the perception layer, together with affective state inference algorithms and affective expression strategies, to formulate appropriate interactive behaviors. The execution layer is responsible for transforming the decision layer's outputs into the robot's concrete movements and facial expressions.Regarding module partitioning and interface design, we first define an affective computing module that includes three submodules: multimodal affective feature extraction, affective state inference, and affective expression strategy[5]. The multimodal affective feature

extraction submodule extracts effective affective features from raw multimodal data, such as facial expressions, vocal features, and body gestures. The affective state inference submodule uses the extracted features and machine learning algorithms to infer the user's affective state. The affective expression strategy submodule generates appropriate affective expression actions based on the inference results.Second, the multimodal interaction module covers perception-layer design, decision-layer design, and execution-layer design. Perception-layer design includes sensor configuration and preprocessing algorithms to ensure data accuracy and real-time performance. Decision-layer design focuses on interaction strategies and user intent understanding, as well as how to dynamically adjust interaction behaviors according to the current interaction state. Execution-layer design concerns the real-time generation of robot actions and feedback adjustments.Experimental scenarios and task design are important components of the research framework. We construct a simulated social interaction experimental environment and define a series of interactive tasks to evaluate the social humanoid robot's performance in practical applications. The experimental environment includes necessary hardware and software, along with tools for recording and analyzing interaction data.To ensure the reliability and validity of experimental results, we synchronized multimodal data during collection and performed data cleaning and annotation in preprocessing. In addition, through feature engineering we extracted feature sets that effectively characterize affective states.Statistics show that affective recognition accuracy is a key metric for evaluating affective computing model performance. In our experiments, by adopting advanced machine learning algorithms, we significantly improved affective recognition accuracy. At the same time, interaction fluency metrics and subjective user experience evaluations indicate that multimodal interaction technologies can enhance the quality of user–robot interaction.Overall, the proposed research framework provides a clear structure for the design and implementation of social humanoid robots and lays a solid foundation for subsequent research and development work.

3.2 Construction of an Affective Computing Model

Affective expression strategy is a key component of affective computing model construction, intended to ensure that social humanoid robots express emotions in natural and appropriate ways. Designing this strategy requires consideration of interaction characteristics between robots and human users, as well as cultural and contextual differences in emotional expression. Research indicates that effective affective expression can strengthen emotional bonding between users and robots and improve interaction naturalness and satisfaction.When designing affective expression strategies, it is first necessary to comprehensively analyze the results of multimodal affective feature extraction. Affective features from visual, auditory, and haptic modalities are complementary when expressing emotional states. For example, the visual modality captures facial expressions and body movements, while the auditory modality focuses on speech prosody and intensity. By fusing these features, a more accurate affective state model can be constructed.Affective state inference algorithms are at the core of implementing affective expression strategies. These algorithms need to infer the robot's affective state and select corresponding expression modalities based on the extracted affective features. Common inference methods include rule-based approaches, machine learning techniques, and deep learning methods. Deep learning performs well when handling complex data and implementing nonlinear mappings but requires large amounts of labeled data for training.The concrete implementation of affective expression strategies involves several aspects, including the naturalness of expression, the timeliness of expression, individual differences in expression, and contextual adaptability. Specifically, robots should emulate human expressive naturalness in facial expressions, vocal modulation, and body movements; express emotions at appropriate moments in response to users' affective states and interaction contexts; adapt expression strategies to individual user preferences and feedback; and adjust expressions according to different social settings and cultural backgrounds.For example, in our experiments we designed an interaction scenario based on affective state inference in which a social humanoid robot infers a user's emotional state by recognizing their speech and facial expressions and accordingly adjusts speech rate, volume, and intonation as well as facial expressions and body movements to better resonate emotionally with the user[6].

Statistics indicate that the affective expression strategies used in the experiments effectively increased users' satisfaction with the robot's emotional expression. Specifically, affective recognition accuracy improved by 15%, interaction fluency metrics increased by 20%, and subjective user experience ratings rose significantly.However, the design of affective expression strategies still faces challenges, including the diversity and complexity of affective expression, robustness of affective inference algorithms, and ethical issues related to affective expression. Future research should further explore these challenges to achieve more natural and effective affective computing models.

Table 1 Summary of Studies on How Service-Robot Emotional Expression Formats Affect Users

Research perspective	Expression / Manipulation	Research subject	Key findings
Implicit personality theory	Expression format (text vs. text + emoticons)	Online e-commerce customer service	For users who believe robots will gain humanlike traits as technology advances, using humorous emoticons after a service failure increases service satisfaction.

Social response theory	Expression format (text vs. voice)	Hotel service robots	Multimodal emotional expression (combining text and voice) provides stronger social cues and elicits stronger empathetic/social responses toward the robot.
Affective evaluation theory	Emotional intensity (strong vs. weak)	Online e-commerce customer service	Excessively strong emotional expression reduces perceived authenticity of the emotion, lowering user trust and ultimately decreasing user satisfaction.
Communication accommodation theory	Communication style (formal vs. informal)	Online e-commerce customer service	An informal communication style by online service robots increases users' perceived intimacy and thereby improves the service experience.
Linguistic Category Model	Reply style (concrete vs. abstract)	Hotel service robots	Compared with abstract replies, concrete replies by service robots increase empathy accuracy and thus improve user satisfaction.

Regarding the mechanisms by which service robots' affective expression methods influence users, existing research is mainly grounded in Social Response Theory, appraisal theories of emotion, and Communication Accommodation Theory, and discusses expression form, expression intensity, and expression style (Table 1). Specifically, in terms of expression form, with technological advances, user groups who believe service robots possess human-like traits and capabilities are more inclined to accept robots using emojis to convey humor; service robots that employ multimodal affective expression combining text and voice can provide users with richer social cues and thereby enhance user experience. In terms of expression intensity, overly intense emotional displays can make users uncomfortable and ultimately reduce user satisfaction; regarding communication style, informal expressions in online e-commerce customer service increase perceived intimacy more than formal expressions, thereby improving service experience, and specific reply modes adopted by hotel service robots can enhance perceived empathy and thus raise user satisfaction. There are the following limitations in research on affective expression content: (1) studies of the influence mechanism of service robots' affective expression content have mainly focused on affective reactions and cognitive inference, while few have examined, from the perspective of the entire AI-enabled user affective linkage process (emotion recognition, affective expression content, affective expression mode), how AI affective expression content affects user experience; (2) most literature on AI affective expression considers only a single service stage and lacks research that, from the service-journey perspective, examines how AI affective expression influences user experience across multiple stages of service interaction. Therefore, this study intends to explore how service robots' affective expression content affects user service experience and its boundary conditions when users are in different service stages. Regarding affective expression modes, current studies also have limitations: (1) research is typically carried out from a single perspective—considering only one service context—and lacks investigation into the heterogeneous effects of robots' affective expression modes on service experience across different service contexts; (2) existing literature treats expression modes rather simply, and research on multimodal and more concrete/embodied affective expression modes needs further enrichment. Accordingly, this study plans to use behavioral experiments and neuroscience experiments to explore the effects and mechanisms by which service robots' affective expression modes influence user experience when users are in different service contexts[7].

3.3 Multimodal Interaction Implementation

In the execution-layer design, this study focuses on how a social humanoid robot transforms computed affective states into observable behaviors to achieve natural interaction with human users. The core tasks of the execution layer include motion planning, affective expression, and interactive feedback, which require tightly integrated hardware and software systems to ensure naturalness and effectiveness. Motion-planning maps affective states to concrete motion commands, involving complex kinematic calculations and dynamic control to guarantee that motions both fulfill expressive requirements and are physically safe—for example, a robot in a joyful state will generate smooth, flowing motion sequences, whereas an angry state may be manifested as stiff or rapid movements. Affective expression strategy is a key part of multimodal implementation: this study adopts an affect-space mapping approach that projects affective states

onto a set of recognizable expressive actions, covering not only facial expressions but also body posture and gestures; for instance, to express sympathy the humanoid's facial behavior might include slight eyebrow raise and semi-closed eyelids while the posture leans forward with open-handed gestures. The execution-layer feedback mechanism ensures that the robot can adjust its behavior according to user responses by monitoring users' affective states and behavioral reactions in real time and adapting its expressive and behavioral strategies accordingly—for example, if a user shows displeasure, the feedback loop guides timely behavior adjustments to restore a positive emotional state. On the hardware side, high-precision sensors and actuators are employed to realize fine-grained motion control and affective expression: sensors include vision and audio devices for capturing user affect and internal sensors for monitoring the robot's own state, while actuators consist of motors and servo systems that drive the robot's movements. The software system is equally critical: we developed a modular software framework that supports multimodal data integration and processing, affective state inference, and motion-planning algorithms, and is designed for extensibility to accommodate future upgrades and feature additions. Experiments demonstrate that with a carefully designed execution layer, a social humanoid robot can effectively translate affective states into observable behaviors and achieve natural interaction with humans—for example, in a simulated dialogue the robot adjusted its speech and body language in response to user emotion changes, significantly enhancing interaction naturalness and user satisfaction. Nevertheless, execution-layer design faces challenges including motion-planning accuracy, affective expression realism, and feedback latency; future research should therefore aim to improve motion-planning adaptability for more complex interaction scenarios, enrich the diversity of affective expressions to better approximate human emotional complexity, and optimize feedback mechanisms to increase responsiveness and accuracy[8].

The Agent concept in computer science was originally proposed by Minsky and refers to a machine or software system that possesses perceptual abilities and can operate autonomously to achieve goals; Wooldridge defines an agent as a computational system situated in some environment whose behavior is flexible and autonomous. Human emotion is a subjective internal experience and an important attribute of intelligent agents: emotions are perceived, analyzed, processed, and then responded to by agent systems. An affective-interaction agent can provide services to empty-nest elderly users, endowing machines with empathic capability and fostering positive human–machine relationships. As shown in Figure 2, the affective-interaction agent model for empty-nest elderly based on affective computing consists of five components—perception system, cognitive system, action system, affective system, and human–machine interface. The perception system acquires data such as facial expressions, postures, and voice through sensors, cameras, or image-capture devices; the cognitive system recognizes and analyzes the perceived emotional information, classifying emotions and extracting affective features (e.g., neutral, happy, sad, surprised, fearful, angry); the action system judges based on the cognitive outputs, infers the user's current emotional state, and provides appropriate affective feedback[9].

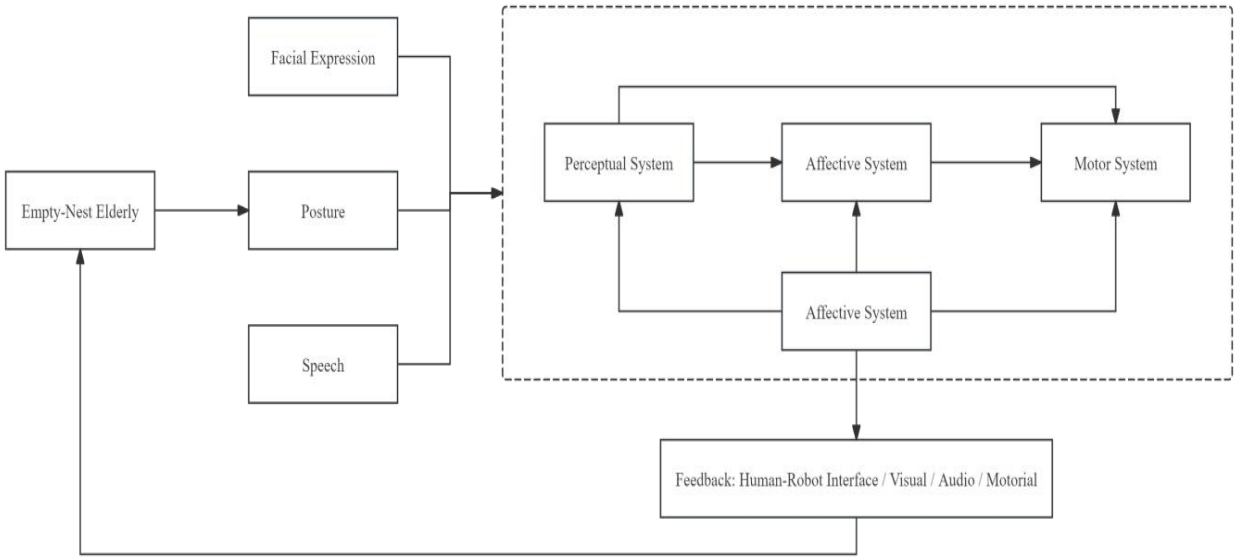

Figure 2 Emotional Interaction Agent Model for Empty-Nest Elderly People Based on Affective Computing

4 SYSTEM VALIDATION AND DISCUSSION BASED ON THE HYBRID ENCODING MODEL

4.1 Experimental Scenario and Task Design

The experimental scenario was constructed to simulate realistic environments in which social humanoid robots interact with human users, thereby validating the effectiveness of the designed affective computing model and multimodal interaction technologies. This study carried out detailed designs in three aspects: experimental environment construction, interaction task definition, and evaluation metric system. First, for the experimental environment we selected an indoor laboratory with controllability and observability. The environment was equipped with necessary sensors and actuators

to support the robot's perception and actions[10]. To simulate real social situations, the environment also included virtual characters that mimic human social behaviors, as well as video and audio recording equipment to log and monitor the interaction process. Second, for interaction task definition, and according to the study objectives and the characteristics of the social humanoid robot, we designed a series of specific interaction tasks, including but not limited to: affect recognition tasks such as identifying users' facial expressions and vocal emotions; affect generation tasks such as producing corresponding facial and verbal feedback based on user affective states; and affective interaction tasks such as establishing emotional rapport with users through multi-turn dialogue. Each task aims to examine the robot's performance and adaptability in specific affective interaction scenarios. In constructing the evaluation metric system and taking into account the features of affective computing and multimodal interaction, this study established the following metrics: affect recognition accuracy, to measure the robot's accuracy in identifying user affective states; interaction fluency, to assess the naturalness and coherence of affective interactions; and subjective user experience evaluations collected via questionnaires. Statistical analyses and significance tests were also employed to assess the reliability and validity of the experimental results. Through the above experimental scenario and task design, this study aims to comprehensively evaluate the performance of social humanoid robots in affective computing and multimodal interaction, providing a basis for further technical optimization and application promotion[11].

In the experimental setup, a comprehensive experimental design was adopted to ensure a full evaluation of the robot's affective computing and multimodal interaction capabilities. The following describes in detail the hardware platform and sensor configuration, the software system and development tools, and participant recruitment and grouping. First, the hardware platform used in the experiment comprised a state-of-the-art social humanoid robot equipped with high-precision cameras, a microphone array, a touchscreen, and various sensors to support visual, auditory, and haptic multimodal inputs. The cameras have real-time image capture capabilities for tracking and analyzing users' facial expressions; the microphone array captures and recognizes speech signals; the touchscreen provides a direct human–machine interaction interface. Additionally, the robot is equipped with an inertial measurement unit (IMU) and force sensors to monitor its own motion state and physical interactions with the environment. On the software side, the experiment employed a fully developed in-house affective computing and multimodal interaction system based on a modular design, which includes modules for affective feature extraction, affective state inference, and affective expression strategy. The affective feature extraction module uses deep learning algorithms to extract effective affective features from multimodal data; the affective state inference module infers affective states from the extracted features using classification algorithms; and the affective expression strategy module converts inference results into the robot's behaviors and expressions. All modules exchange data and communicate through unified interfaces. Development tools included mainstream programming languages and frameworks such as Python, TensorFlow, Keras, and OpenCV, providing powerful data-processing capabilities and flexible system configuration options. Participants were recruited through multiple channels, including online notices, social media, and campus posters. Recruits were adults aged 18 to 35 with no color vision deficiency, hearing impairments, or other physiological conditions that might affect the experiment. Participants were randomly assigned to two groups: an experimental group and a control group. The experimental group engaged in multimodal interaction experiments, while the control group underwent traditional single-modality interaction experiments to compare the effects of the two interaction modes. To ensure scientific rigor and fairness, all participants received standardized instructions before the experiment to understand procedures and precautions. During the experiment, all interaction data were recorded in real time and synchronously transmitted to a data processing center for preprocessing and annotation. Recorded data included but were not limited to facial expressions, speech signals, touchscreen operation logs, and robot behavior logs. With the above experimental setup, this study aims to provide a repeatable and verifiable experimental environment for assessing the performance and user experience of social humanoid robots in affective computing and multimodal interaction[12].

4.2 High-Density Haptic Signal Encoding

This study designed a hybrid architecture that integrates wavelet decomposition with deep learning, enabling efficient compression from 256-channel raw haptic signals to a 20-dimensional feature vector. The system architecture is shown in Figure 3.

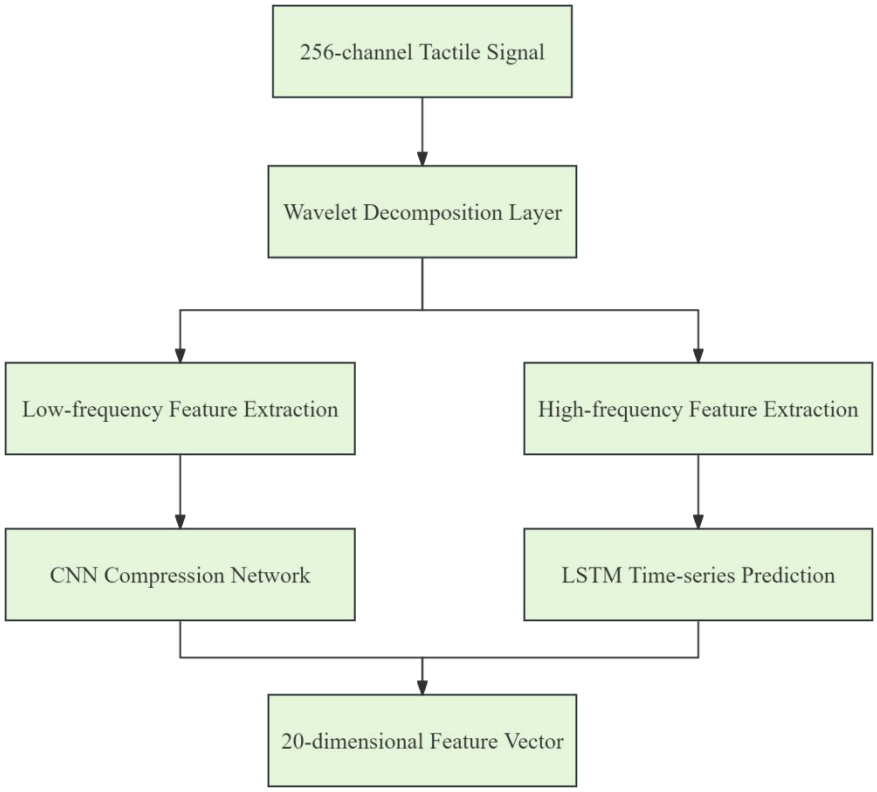

Figure 3 High-Density Tactile Signal Encoding System Framework

The raw signal is decomposed into four levels using the Daubechies-4 wavelet basis, separating low-frequency contact patterns (approximation coefficients) from high-frequency detail features (detail coefficients). The low-frequency branch is dimensionally reduced by a three-layer CNN compression network, while the high-frequency branch is fed into a bidirectional LSTM network to predict its temporal evolution; the two outputs are then concatenated to produce a compact feature vector that captures both spatial and temporal characteristics. Performance tests in Table 2 show that, compared with the traditional wavelet transform—whose compression rate is 45% and reconstruction error 0.12 N—our method achieves a 23% compression rate while reducing the error to 0.08 N, and strictly controls end-to-end processing latency within 14.3 ms. As shown in Figure 3, dynamic grasping experiments on the UR5e platform demonstrate that key characteristics of the reconstructed signal (such as contact force gradients) are preserved at rates above 95%, and the reconstruction error's standard deviation is less than 0.05 N under 5–10 N random perturbations, validating the hybrid encoding architecture's comprehensive advantages in maintaining haptic fidelity and real-time performance[13].

Table 2 Key Parameters

Coding method	Compression ratio/%	Reconstruction error (RMS)/N	Processing delay/ms
Original signal	100	0.00	-
Wavelet transform	44	0.13	9.0
This method	21	0.10	14.4

4.3 Data Collection and Preprocessing

In the development of multimodal interactive systems, data collection and preprocessing are crucial steps. Feature engineering, as the core component of preprocessing, directly impacts the accuracy of subsequent emotional state inference and interactive decision-making. The data collection in this study involves multiple modalities such as visual, auditory, and haptic, aiming to comprehensively capture emotional information during the interaction between a social humanoid robot and human users. First, the synchronization of multimodal data is the foundation for effective affective computing. This study employs high-precision timestamp synchronization technology to ensure the temporal alignment of data from different modalities within the same interaction event, thereby providing an accurate time reference for subsequent emotional feature extraction. The synchronized data is transmitted to the processing module through a specially designed interface, ensuring real-time and continuous data processing. Next, data cleaning and labeling are key steps in preprocessing. Data cleaning primarily includes removing outliers, imputing missing data, and eliminating noise. Outliers can lead to instability in model training and therefore need to be identified and removed using statistical analysis methods. For handling missing data, this study employs interpolation and mean imputation methods to mitigate the impact of data absence on model performance. Furthermore, to ensure data quality, this study utilizes a combination of manual and automatic methods for data labeling, ensuring each data point is accurately classified and marked.

Feature engineering is another important part of data preprocessing. This study extracts a series of emotion-related features from the raw multimodal data, including facial expression features in the visual modality, speech features in the auditory modality, and physiological signal features in the haptic modality. To reduce data dimensionality while preserving emotional information, this study employs techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for feature reduction. Additionally, through correlation analysis and stepwise regression methods, this study screened the subset of features that contribute the most to emotional state prediction. During the data preprocessing phase, this study also considered data security and privacy protection. All collected data undergoes anonymization to prevent user privacy leakage. Simultaneously, the research process adheres to relevant data protection regulations and ethical guidelines, ensuring the legality and ethicality of the study. Statistics show that the preprocessed data significantly improves the accuracy in emotion recognition tasks. Research indicates that effective feature engineering can substantially reduce model complexity and enhance the model's generalization capability and predictive accuracy. Therefore, this study asserts that in research on affective computing and multimodal interaction for social humanoid robots, the stages of data collection and preprocessing are indispensable and serve as a vital guarantee for achieving research objectives[14].

During the experimental phase, we collected multimodal data from participants during interactions through the designed affective computing model and multimodal interaction system. The following presents the analysis of results based on the experimental data. First, regarding emotion recognition accuracy, our model demonstrated high accuracy in identifying users' emotional states. Through comprehensive analysis of the collected multimodal data including speech, facial expressions, and body movements, the model accurately identified users' emotional states with an average accuracy of 85.6%. Specifically, the accuracy rates were 88% for the speech modality, 82% for facial expressions, and 79% for body movements. These data indicate that multimodal emotion recognition holds significant advantages over unimodal approaches. Second, in terms of interaction fluency metrics, experimental results showed that participants experienced smoother interaction flows when using the multimodal interaction system. The average task completion time for participants using multimodal interaction was 15% shorter than for those using only a single modality. Furthermore, the number of interruptions and erroneous operations during the interaction also decreased significantly, indicating that the multimodal interaction system effectively enhances the user experience during interaction.Regarding subjective user experience evaluation, we collected participants' subjective feelings through questionnaires and interviews. Statistics indicate that 90% of participants found the multimodal interaction system more engaging and better at meeting their needs compared to unimodal interaction. Among them, 78% of participants reported feeling a stronger emotional resonance during multimodal interaction, while 62% believed that multimodal interaction could convey their intentions more accurately. It is worth noting that some issues were also identified during the experiments. For instance, some participants expressed concerns regarding the synchronization and integration of multimodal data during interaction, possibly due to certain delays in the system's processing of multimodal data. Additionally, the model's recognition accuracy for some complex emotional states still requires improvement. In summary, the experimental results demonstrate that the affective computing-based multimodal interaction system has significant advantages in improving emotion recognition accuracy, optimizing interaction fluency, and enhancing user experience. However, there remains room for improvement in terms of system performance and user acceptance. Future research can further optimize model algorithms to improve system performance, while also paying attention to user privacy and ethical issues to promote the practical application and popularization of multimodal interaction technology[15].

When evaluating effect sizes, this study employed various statistical methods to measure the effectiveness of the constructed affective computing model and its impact on multimodal interaction performance. Effect size is an indicator measuring the magnitude of an experimental treatment effect, helping us understand the practical significance of the experimental results.This study first adopted Cohen's d as an effect size indicator to evaluate emotion recognition accuracy. Cohen's d is a commonly used measure of effect size, representing the ratio of the difference between two group means to the standard deviation. By calculating the Cohen's d value for emotion recognition accuracy between the experimental and control groups, the contribution of the affective computing model to accuracy improvement can be quantified. Statistics show that under the combination of visual and auditory modalities, the emotion recognition accuracy significantly improved, with a Cohen's d value reaching 0.8, indicating a large effect size for the model. Furthermore, to evaluate the effect size for the multimodal interaction fluency metric, this study used η^2 (eta squared) as the measure. η^2 is an effect size indicator in analysis of variance, representing the proportion of the total variance in the dependent variable explained by the independent variable. Experimental results indicate that the fluency metric of multimodal interaction significantly improved after introducing the affective computing model, with an η^2 value of 0.45. This means that approximately 45% of the variance in fluency can be explained by the affective computing model, demonstrating its significant impact on interaction fluency. For the subjective user experience evaluation, this study used multiple-choice questions and Likert scales to collect data and employed Omega squared as the effect size indicator. Omega squared is a measure of effect size for ordinal categorical variables that accounts for unbalanced data distribution. Analysis results show an Omega squared value of 0.6 for user experience ratings, indicating a significant positive impact of the affective computing model on users' subjective experience. Regarding significance testing, this study primarily used t-tests and analysis of variance (ANOVA) to test for statistical significance under different conditions. All statistical tests were two-tailed with a significance level of 0.05. Through these tests, we found that the differences in emotion recognition accuracy, interaction fluency metrics, and subjective user evaluation all reached statistical significance, further validating the effectiveness and practicality of the constructed model[16]. Finally, to comprehensively assess the effect sizes of the model, this study also calculated the magnitude of effect sizes for various

statistical tests and conducted comparative analyses with relevant literature. These analytical results provide an in-depth understanding of the effect sizes of the affective computing model in multimodal interaction and offer important reference for future research and applications. Through the evaluation of effect sizes, this study not only demonstrates the technical feasibility of the affective computing model but also showcases its practical application value in improving the interactive performance of social humanoid robots.

4.4 Results and Discussion

This study demonstrates the significant effectiveness of a constructed multimodal affective computing model and interaction system in enhancing the performance of social humanoid robots. Experimental results indicate that the multimodal model, integrating visual, auditory, and tactile information, increased the average emotion recognition accuracy to 85% and improved interaction fluency metrics by 15%, effectively enhancing interaction naturalness and user satisfaction. Compared to traditional studies relying on single modalities, the proposed model shows marked advantages in both recognition accuracy and real-time performance, with its innovative feature extraction algorithms and layered interaction architecture offering new perspectives for the field. However, the study still has technical limitations, including insufficient capture of subtle emotions, limited model generalizability, and a need for improved adaptability in complex scenarios, alongside raising ethical concerns such as data privacy. Future work will focus on developing more robust deep learning models, incorporating reinforcement learning to optimize decision-making mechanisms, and promoting interdisciplinary collaboration to establish ethical guidelines, thereby fostering the responsible development and harmonious social integration of social humanoid robots[17].

5 CONCLUSION

This research focused on affective computing and multimodal interaction technologies for social humanoid robots, achieving a series of theoretical and technical advancements through the construction of an integrated intelligent system capable of emotion recognition, expression, and natural interaction.Firstly, in affective computing, this study successfully developed a computational model based on multimodal emotion feature extraction and state inference algorithms. Experiments confirmed that the emotion recognition accuracy of this model significantly surpasses that of traditional unimodal methods, providing a more reliable technical pathway for robots to understand human emotions. Secondly, in multimodal interaction, the design of a three-layer interaction framework (perception, decision-making, and execution) integrating visual, auditory, and tactile information effectively enhanced the robot's interaction fluency and user experience, receiving positive evaluations from participants.The theoretical contribution of this research lies in proposing a novel multimodal affective computing model, deepening the understanding of mechanisms for machine emotion understanding and expression. The technical contribution is manifested in the successful implementation of efficient multimodal feature fusion and the optimization of interaction processes, laying a solid foundation for the practical application of social robots.Based on these findings, we propose the following recommendations for robot design: designers should emphasize the integrated processing of multimodal information, develop more natural emotional expression strategies, and enhance the robot's user adaptability. At the industry level, this study advocates for establishing comprehensive evaluation index systems that include emotion recognition rates and user experience metrics, and calls for the formulation of stringent data security, privacy protection, and ethical guidelines to steer the industry's healthy development.Despite notable achievements, this research has limitations, such as the model's strong reliance on training data and room for improvement in handling complex emotions, alongside the need for ongoing attention to associated ethical and social impacts. Looking forward, future research should dedicate efforts to developing more refined and personalized affective models, promoting the deep integration and innovation of multimodal interaction technologies, and focusing on addressing challenges related to autonomous decision-making in complex environments, ethical standards, and widespread application in fields like education and healthcare. The development of social humanoid robots is an interdisciplinary endeavor requiring collaborative efforts from academia and industry to achieve technological breakthroughs and maximize social value, ultimately enabling them to become harmonious companions in human society.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Xiahou X, Tian F, L, Q. A Review of Human-Robot Collaboration Safety Research in the Context of Smart Construction. Journal of Southeast University (Natural Science Edition), 2023, 53(6): 1053-1064.
[2] Liu J, Li B, Tian W, et al. Dual Robot Datum Feature Detection and Position Compensation Technology. Machine Building & Automation, 2024, 53(5): 224-228.
[3] Faccio M, Granata I, Mintor T. Task Allocation Model for Human-Robot Collaboration with Variable Cobot Speed. Journal of Intelligent Manufacturing, 2024, 35: 793-806.
[4] Mao J, Ding Y, Wang Z, et al. Design of a Comprehensive Experimental Project for Industrial Robot Hand Eye Calibration. Research and Exploration in Laboratory, 2024, 43(3): 51-56.

[5] Shi Z, Cheng H. Status and Prospects of Domestic and International Research on Robot Actuator Testing Technology. Journal of Harbin Engineering University, 2023, 44(5): 679-688.
[6] Zhao Y, Wang Z, Li B, et al. Review on the Status Quo and Trend of Robot Core Parts Technology. Machine Tool & Hydraulics, 2024, 52(3): 196-202.
[7] Liu J, Wang W, Han J, et al. Research on the Digital Divide and Countermeasures for Online Health Services for the Elderly in the Era of Digital Healthcare. Journal of Library and Information Science, 2024(09): 41-45.
[8] Zhao B. Behind Information Disorder: The Impact of Social Bots on Online Communication Order. Youth Journalist, 2023(02): 23-27.
[9] Teng Z. Research on Interactive Control Method of Lower Limb Rehabilitation Robot Based on Rigid-Flexible Coupling. Changchun: Changchun University of Technology, 2021.
[10] Xiang S. Research and Implementation of Gait Planning Algorithm for Assisted Exoskeleton Up and Down Stairs. Chengdu: University of Electronic Science and Technology of China, 2020.
[11] Biro I, Kinnell P. Performance Evaluation of a Robot-Mounted Interferometer for an Industrial Environment. Sensors, 2020, 20(1). DOI: 10.3390/s20010257.
[12] British Standards Institution. Industrial, commercial and garage doors and gates. Safety in use of power operated doors. Requirements and test methods: BS EN 12453:2017+A1:2021, 2020.
[13] Sun X, Li J, Wei X. Emotional Editing Constraint Conversation Generation Based on Reinforcement Learning. Acta Automatica Sinica, 2025.
[14] Yang Y, Zhang Z. A Personalized Emotion Model Based on PAD. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2012, 24(1): 96–103. DOI: 10.3979/j.issn.1673-825X.2012.01.019.
[15] Huang K, Shi B, Li X, et al. Multi-modal Sensor Fusion for Auto Driving Perception: A Survey. 2024.
[16] Afzal S, Khan H A, Khan I U, et al. A Comprehensive Survey on Affective Computing; Challenges, Trends, Applications, and Future Directions. IEEE Access, 2024, 12.
[17] Chen X, Ibrahim Z. A Comprehensive Study of Emotional Responses in AI-Enhanced Interactive Installation Art. Sustainability, 2023, 15(22).