This is a peer-reviewed, accepted author manuscript of the following conference paper: Ait Ameur, MA et al., 2025, MICA: a multimodal intelligent cognitive assessment framework integrating generative AI and social robot for early cognitive intervention. in D Cafolla, T Rittman & H Ni (eds), Artificial Intelligence in Healthcare: Second International Conference, AIiH 2025, Cambridge, UK, September 8–10, 2025, Proceedings, Part I. Springer, Cham, pp. 96-109. https://doi.org/10.1007/978-3-032-00652-3_8

# MICA: A Multimodal Intelligent Cognitive Assessment Framework Integrating Generative AI and Social Robot for Early Cognitive Intervention

Mohamed Adlan Ait Ameur[1], Erfu Yang[1]*, William J. McGeown[2], and Yin-Ping Zhang[3]

[1] Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Glasgow, G1 1XJ, United Kingdom
[2] Department of Psychological Sciences and Health, University of Strathclyde, Glasgow, G1 1QE, United Kingdom
[3] Faculty of Nursing, Xi'an Jiaotong University, Health Science Center, No.76, West Yanta Road, Xi'an, Shaanxi, 710061, P.R. China
adlan.ait-ameur@strath.ac.uk, erfu.yang@strath.ac.uk, william.mcgeown@strath.ac.uk, cathyzh@mail.xjtu.edu.cn

**Abstract.** The rapidly growing older adult population underscores the urgent need for innovative solutions to detect, classify, and monitor early cognitive decline. Traditional cognitive screening methods, such as paper-and-pencil tests like the Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE), suffer from notable limitations, including low patient engagement, limited motivational appeal, inadequate sensitivity to subtle cognitive decline, and susceptibility to practice effects with repeated administrations. Furthermore, such tests require substantial clinician time, as they must be administered by trained healthcare practitioners, increasing healthcare costs. Given the current shortage of effective interventions and reliable screening methods for early-stage cognitive decline, socially assistive robots offer a promising dual function: They can deliver personalized and engaging cognitive stimulation and social support while monitoring cognitive health. This paper proposes a new Multimodal Intelligent Cognitive Assessment (MICA) framework integrated into the Pepper social robot and enhanced by generative AI technologies. MICA consists of three core components:(1) a conversational and cognitive exercise interface powered by generative AI, enabling natural, engaging interactions tailored to various cognitive domains; (2) multimodal perception capabilities, incorporating emotion recognition using DeepFace, robust speech recognition, and real-time personalized adaptation based on emotional and cognitive feedback, and (3) an advanced performance logging system designed to systematically record patient accuracy, response time, and emotional states. Initial evaluations with Pepper demonstrated real-time emotion detection and adaptive exercises, which illustrate the potential for high levels of engagement and early intervention. Although current evaluations were conducted by the authors, more comprehensive user studies are planned to validate the effectiveness of MICA within the populations of interest.

**Keywords:** Generative AI, Assistive Robotics, Cognitive Assessment, Social Robotics, Human-Robot Interaction (HRI)

* Corresponding author: erfu.yang@strath.ac.uk

## 1 Introduction

According to the World Health Organization, population ageing is accelerating rapidly. The number of people aged 60 and older is projected to increase from 1 billion in 2020 to 1.4 billion by 2030 and 2.1 billion by 2050. This trend highlights a significant demographic shift, with the older adult population growing much faster than the general population [1]. Ageing is inherently associated with declining health, increasing the demand for long-term care services [2]. A key challenge accompanying ageing is cognitive decline, often progressing into neurodegenerative conditions such as Alzheimer's disease. According to the World Health Organization (WHO), dementia currently affects more than 55 million people globally, with Alzheimer's disease accounting for approximately 60–70% of cases. Nearly 10 million new dementia cases emerge annually, primarily in low- and middle-income countries. Dementia is also the seventh leading cause of death globally and a major source of disability and dependency among older adults, imposing immense economic and social burdens [3]. In 2019 alone, dementia-related costs reached approximately US$1.3 trillion worldwide, with around half attributable to informal care provided by family members or close friends, who typically dedicate about five hours per day to caregiving activities [3]. These conditions significantly impact individuals' self-care abilities and escalate the necessity for medical and social support [4]. Furthermore, existing treatment options for Alzheimer's disease and related neurodegenerative disorders remain limited, complicated by issues such as drug tolerance and adverse medication reactions, which underscores the urgency for innovative preventive and therapeutic solutions [2].

Social-assistive robots (SAR) have emerged as practical tools to provide personalized cognitive support, improve patient well-being, and optimize healthcare efficiency [5]. Integrating social robots in healthcare offers a promising solution to address workforce shortages and improve care accessibility [6]. Intelligent robotic systems have the potential to transform diagnosis, treatment, and patient care in healthcare settings [7]. Humanoid and companion robots, including NAO and PARO, have shown promising results in improving cognitive performance and promoting social and emotional well-being among older adults [8]. When being integrated with artificial intelligence (AI), including deep learning technologies, these robots hold strong potential to support individuals experiencing age-related cognitive decline effectively. Furthermore, social robots provide a practical solution to alleviate the economic strain on healthcare systems while supporting independent living for older populations [9].

Personalization in care is critical because cognitive decline and emotional needs vary considerably among individuals. SAR systems are distinguished by their ability to offer personalized patient interactions [10]. Personalized interactions through social robots can significantly enhance patient engagement and improve cognitive outcomes [11]. Effective robotic integration, therefore, requires careful customization of interactions, aligning robot functionality, motivational strategies, and communication methods to individual patient needs. Successful

robotic systems must also dynamically adapt to patient behaviour during interactions [12].

Early detection of cognitive decline is essential to implement timely interventions and potentially delay the progression of cognitive impairments. Emotional monitoring has the potential to play a larger role in the early detection of cognitive decline and in providing personalized support, as emotional states significantly influence cognitive performance and patient engagement. This is supported by evidence showing that emotion modulates attentional selectivity and executive functions, which subsequently enhance learning and memory processes [13]. Incorporating emotional feedback via multimodal analysis, such as facial expression recognition with technologies such as DeepFace [14], may also provide more personalized and adaptive cognitive interventions.

Memory, a core cognitive function, tends to deteriorate as people age [15], with more substantial impairment tending to be a key indicator of dementias such as Alzheimer's disease. Working memory, which allows temporary storage and real-time information processing, is crucial in supporting complex cognitive activities such as language, reasoning, and problem-solving [16]. The long-term memory system, on the other hand, enables the storage of memories that relate to episodic and semantic information over extended periods of time. Cognitive tasks that are designed to enhance these memory systems are often perceived as repetitive and frustrating, leading to low engagement and reduced data quality. Using gamification features can enhance engagement, improve data quality, and increase intervention effectiveness without compromising scientific validity [17]. Socially assistive robots such as Pepper provide a friendly look and can deliver personalized games by responding to patients' mental and emotional states.

Despite these advancements, there remain notable gaps in current research, particularly regarding personalization, multimodal information capture for further analytics, and integrating feedback based on emotional state into cognitive exercises. To address these gaps, the primary aim of this system is to improve patient engagement through personalized cognitive assessment and training. Specifically, the objectives of this study include leveraging generative AI capabilities to simplify and streamline cognitive game development, reducing reliance on manual programming, and fostering more responsive and individualized interventions. The proposed robotic system will be able to dynamically generate and adapt cognitive exercises by incorporating large language models (LLMs) like ChatGPT.

The structure of this paper is as follows: Section 2 provides a review of related work. Section 3 outlines the proposed methodology. Section 4 presents the results along with a discussion. Section 5 addresses the limitations and future work of the study, and Section 6 presents the study's conclusion.

## 2 Related Works

Social robots have increasingly demonstrated potential in therapeutic interventions, particularly for older adults experiencing cognitive impairments. Caroppo et al. in [18] explored affective responses among older adults undergoing cognitive stimulation therapy assisted by the Pepper robot. Their findings indicated

that social robots effectively evoke positive emotional reactions; however, these reactions vary significantly according to the severity of cognitive impairment. Individuals experiencing more pronounced cognitive decline demonstrated fewer positive emotional responses, highlighting the critical need for personalized cognitive interventions that adapt to patients' specific cognitive status. Expanding on this idea, [19] conducted a systematic review focused on human-robot interaction (HRI) strategies designed to enhance social engagement in elderly individuals with mild cognitive impairment (MCI). Analyzing 15 distinct studies, the authors identified several influential factors, including robot usability, patient acceptance, engagement quality, and interaction barriers. Their review underscored the promising role of social robots in cognitive therapy while emphasizing personalized and interactive interventions.

The authors of [20] proposed a cognitive therapy system utilizing the social robot EBO, which incorporates large language models (LLMs), automatic speech recognition (ASR), and text-to-speech (TTS) capabilities. Built on the CORTEX architecture, this system integrates a therapist-in-the-loop for session preparation while autonomously administers adaptive cognitive therapies. Despite its autonomy, the effectiveness of personalization in this system heavily depends on the therapist's initial configuration, potentially restricting real-time adaptability during cognitive sessions. Furthermore, research on dynamically adjusting exercises based on real-time patient performance and emotional state is limited, potentially affecting long-term efficacy and responsiveness to evolving patient needs. Addressing the importance of adaptability and safety, [21] introduced a cognitive robotic system tailored explicitly for brain training in individuals with mild dementia. Their system featured adaptive and safety modules, with preliminary laboratory results indicating effective patient interaction and system flexibility. Nonetheless, the authors acknowledged limitations regarding the system's robustness, perceptual capabilities, and safety, emphasizing the necessity of further enhancements before deployment in clinical environments.

Additionally, [22] developed ReMe, an AI-driven chatbot framework designed explicitly for personalized cognitive training. Using large language models, ReMe aims to boost episodic memory and enhance patient engagement through interactive exercises. While promising, further evaluation must confirm its real-time adaptive personalization capacities for diverse cognitive profiles and emotional states. These studies highlight a substantial research gap concerning the adaptive personalization of robotic cognitive stimulation tailored explicitly to individual emotional and cognitive profiles. To address this gap, we propose a novel cognitive exercise framework called MICA (Multimodal Intelligent Cognitive Assessment), as shown in Figure 1, which includes generative AI and interaction capabilities to provide engaging and personalized cognitive exercises in response to patients' cognitive and emotional conditions.

## 3 Methodology

Human-robot interaction (HRI) often depends on effectively integrating diverse communication signals, including speech, visual cues, text, gaze, and touch, collectively known as multimodal interaction. Recent advances in cognitive science,

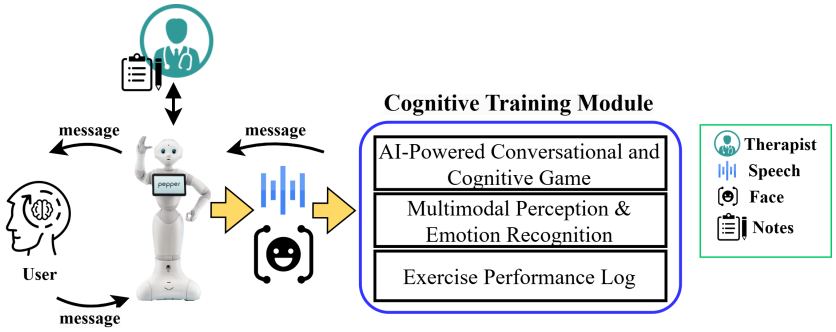


**Fig. 1.** MICA (Multimodal Intelligent Cognitive Assessment) with Pepper robot

ergonomics, and virtual reality have driven substantial progress in multimodal AI applications, particularly in emotional data processing, psychological assessments, remote education, and rehabilitative healthcare [23]. Currently, interactions with large language models (LLMs) have evolved through prompting techniques, which involve the input of textual cues or an initial context to guide model output, enhancing coherence and relevance [24]. Prompt engineering is central to effective interactions with Large Language Models (LLMs), which involves crafting and refining prompts to direct the model toward desired, specific outputs [25].

Leveraging this methodology, we adopt the concept of Prompt Chaining, a type of iterative prompting that enables multi-step reasoning by feeding each output into the subsequent prompt [26]. Within our MICA (Multimodal Intelligent Cognitive Assessment) framework, predefined prompt chains have been developed to guide the execution of cognitive exercises consistently. Activated via voice-based wake words, these prompts enable immediate and intuitive initiation of cognitive tasks such as the Word Association and Word Chain games, each targeting distinct cognitive domains implemented during this initial stage of development. The system provides patients, healthcare practitioners, or carers with simple, natural, and rapid access to these preconfigured cognitive interactions, significantly enhancing usability and ensuring standardized, personalized delivery of cognitive assessments and interventions. The MICA framework, integrated into the Pepper robot, comprises three main components:

### 3.1 AI-Powered Conversational and Cognitive Game component:

The initial component of the MICA framework involves generating natural and cognitively targeted interactions between the patient and the Pepper robot through prompt chains. It comprises two key elements: a general system prompt that facilitates intuitive dialogue and enhances user comfort and a set of task-specific chains tailored to deliver cognitive exercises such as Word Association and Word Chain games. These exercises target critical domains, including attention, memory, executive function, and language skills. All prompts are pre-

defined, minimizing manual programming and ensuring consistent, personalized delivery.

### 3.2 Multimodal Perception and Emotion Recognition component:

The second component provides the critical multimodal perception and adaptive feedback mechanisms central to the MICA framework. It integrates advanced multimodal technologies, including real-time emotion recognition using the DeepFace framework and robust speech recognition setup with Google's Speech Recognition API [27]. Emotion recognition continuously analyzes facial expressions to assess and record emotional states, directly informing Pepper's interaction strategies. Moreover, speech recognition ensures an accurate transcription of patient responses despite environmental acoustic challenges. Combining these multimodal inputs, the system dynamically adapts to real-time interactions, modulating engagement, task difficulty, pacing, and motivational feedback based on cognitive performance and emotional cues. This comprehensive multimodal integration enables personalized, responsive, and engaging cognitive interventions tailored precisely to individual patient needs.

### 3.3 Exercise Performance Logging component:

The third component of the MICA framework is the Exercise Performance Log, designed explicitly to record systematically comprehensive patient-interaction data. This component continuously captures performance-related metrics during each cognitive session, integrating several evaluation dimensions. Specifically, it logs cognitive performance metrics, including accuracy (percentage of correct responses), response time (in milliseconds), frequency, and type of patient errors. Multimodal interaction data from real-time emotional recognition are recorded through the DeepFace framework and speech recognition accuracy metrics, such as word error rate (WER) and confidence scores. The logged data points are structured into a unified JSON or CSV format, enabling real-time personalization and detailed post-session analyses. Clinicians and researchers can leverage these data to monitor patient progress, identify cognitive or emotional trends, and refine future cognitive interventions.

Working memory, the capacity to temporarily hold and mentally manipulate information, is crucial for following instructions, engaging in conversations, and solving problems. To illustrate the potential of MICA and to support and enhance this cognitive domain, as proof of concept, we designed a set of interactive cognitive exercises structured around prompt chains. Specifically, two exercises were implemented: the Word Association game and the Word Chain game. Table 1 summarizes the cognitive domains targeted by each of these activities. The therapist, for example, guides the selection of the appropriate activity based on the individual's cognitive needs and session objectives. During the Word Association game, the robot provides thematically generated word lists, and the patient responds by making associations, thereby engaging targeted cognitive functions. Alternatively, the Word Chain game may be selected, in which a starting word is dynamically generated by ChatGPT, followed by alternating responses from

the patient and the Pepper robot. Each patient response is validated by ensuring it begins with the final letter of Pepper's previous word. The prompt chains for exercises are preconfigured and can be activated via voice commands, allowing for a more natural and accessible patient experience. If it is deemed more appropriate by the health professional, a traditional conversation mode remains available through Pepper's standard prompt-based interaction.

**Table 1.** Cognitive domains targeted by the proposed cognitive exercises

| Cognitive Domain | Word Association Exercise | Word Chain Exercise |
|---|---|---|
| Attention | Yes (attention to stimulus prompt) | Yes (sustained attention to sequence) |
| Language | Yes (lexical retrieval and verbal fluency) | Yes (vocabulary usage and syntactic skills) |
| Executive Function | Yes (categorization, semantic linking) | Yes (rule adherence, strategic planning) |
| Working Memory | Yes (retrieval of previously heard terms) | Yes (maintenance of sequential constraints) |
| Orientation | No | No |
| Visuospatial Skills | No | No |
| Abstract Thinking | Yes (establishing semantic associations) | Yes (reasoning within defined constraints) |

To construct the system architecture, we build upon prior research utilizing ZeroMQ-based messaging and chatbot-driven interactions for real-time human-robot communication with the Pepper robot [28]. Specifically, we adapt and extend this approach to enable seamless integration between Pepper and our cognitive training module powered by ChatGPT (gpt-4-turbo-2024-04-09), ensuring responsive and context-aware cognitive interactions. Our system leverages real-time emotional feedback to enable personalized cognitive exercises, continuously captured using DeepFace and transmitted via ZeroMQ messaging. However, during the interaction, we observed that Pepper's built-in microphones near its internal cooling fans introduced acoustic interference, significantly reducing speech recognition accuracy. To address this issue, we integrated an external microphone alongside an audio-processing pipeline. The external audio input, captured at 48 kHz, is processed through a WebRTC-based Voice Activity Detection (VAD) module, enabling effective speech segmentation and improving the overall performance of our real-time speech recognition system. The segmented speech data is then transcribed into text using Google's Speech Recognition API and transmitted via ZeroMQ sockets to the ChatGPT-based module, ensuring robust and accurate conversational interactions. A dedicated ZeroMQ control socket is also implemented to pause or resume audio capturing as required. Concurrently, DeepFace analyzes images from Pepper's camera to detect real-time emotion. Although it runs continuously, emotion detection is specifically time-locked and segmented to correspond precisely to critical task periods, ensuring relevant emotional responses are captured in context. The detected emotions are communicated via ZeroMQ, dynamically influencing the chatbot-generated instructions and enhancing the responsiveness of human-robot interaction. This enables Pepper to adjust interaction styles and pacing dynamically based on emotional responses. Through this multimodal, emotionally responsive integration, the proposed framework MICA offers an innovative solution for personalized cognitive training, effectively bridging cutting-edge HRI technology with mental healthcare. The system architecture is shown in Figure 2.

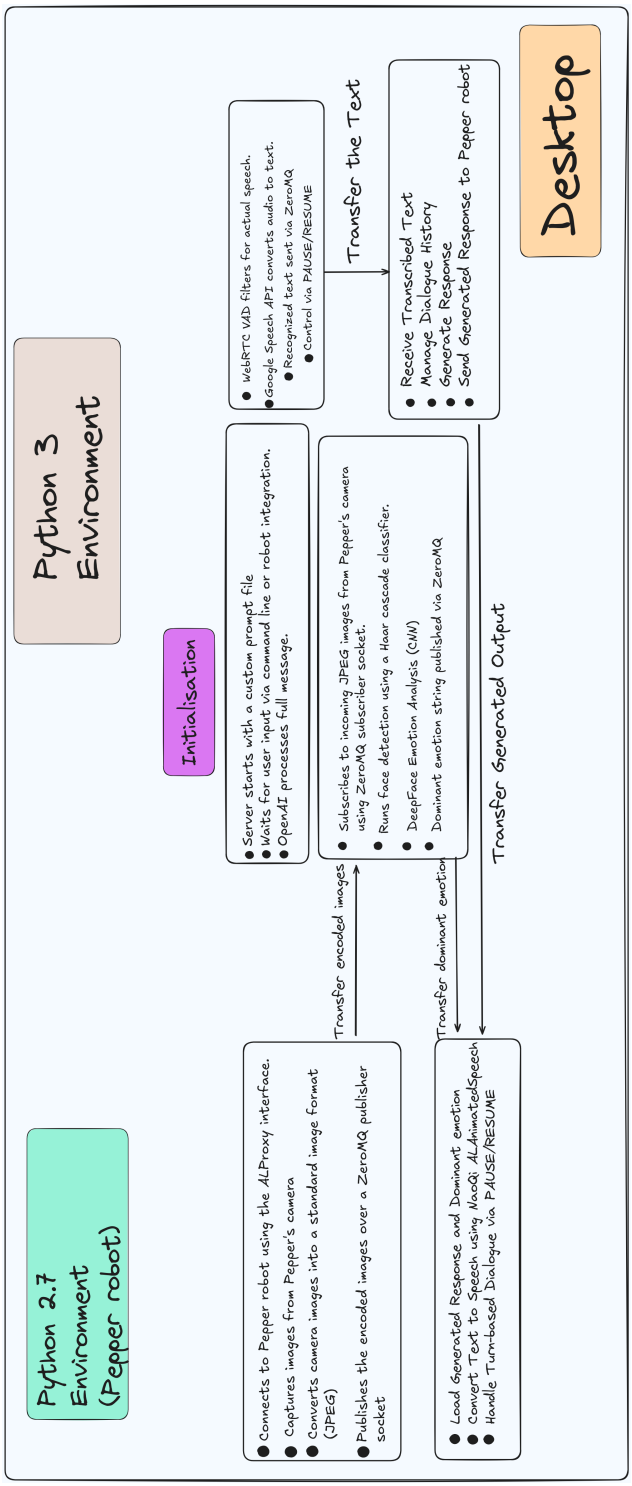


**Fig. 2.** Architecture of the MICA framework: Real-Time emotion and dialogue processing

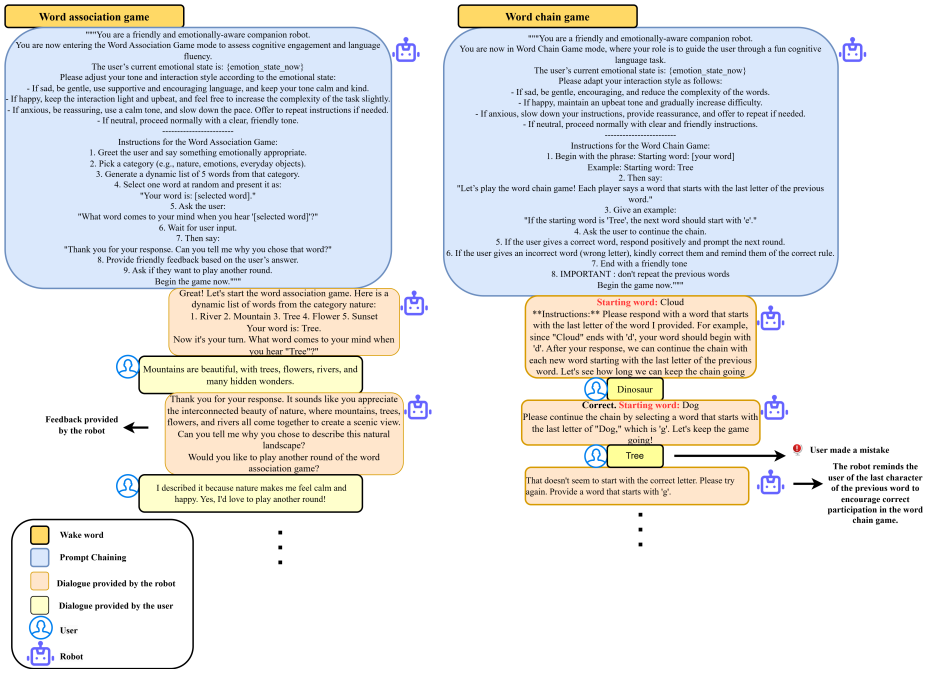


**Fig. 3.** Interactive cognitive training with Pepper robot: Word Association and Word Chain Games

## 4 Results and Discussion

The evaluation involved two cognitive exercise games, the *Word Association Game* and the *Word Chain Game*, implemented and tested through direct interaction with the Pepper robot. In the *Word Association Game*, Pepper dynamically generated a thematic list of words (e.g., nature) and prompted the user to express associated thoughts or feelings. User responses successfully demonstrated cognitive associations and emotional reactions. Pepper provided real-time tailored feedback, enhancing engagement through positive emotional reinforcement, as illustrated in Figure 3. This exercise effectively captured emotional states and cognitive responsiveness, particularly highlighting associations related to positive emotions such as calmness and happiness. The thematic prompts could also be adaptively tailored in response to the user's reactions, allowing the operator or therapist to deliberately foster desired emotional states such as happiness, social engagement, or nostalgia while avoiding negative states like distress or anxiety. The *Word Chain Game* evaluated attention, memory, and cognitive flexibility. The user was required to sequentially provide words starting with the last letter of Pepper's previous word. During testing, occasional mistakes occurred (e.g.,

selecting a word beginning with an incorrect letter). Pepper effectively identified these errors, offering corrective and supportive feedback and reinforcing instructions to keep the user on track.

**Exercise Performance Logging Component**


```
{
  "session_id": "session_001",
  "timestamp": "2025-03-28T10:30:00",
  "task_name": "Word Chain Game",
  "user_responses": [
    {
      "response": "Apple",
      "correct": true,
      "response_time_ms": 1450,
      "speech_recognition_confidence":0.96,
      "emotion_detected": "happy",
      "emotion_confidence": 0.98
    },
    {

      "response": "Elephant",
      "correct": false,
      "error_type":"Initial letter mismatch",
      "response_time_ms":2200,
      "speech_recognition_confidence":0.91,
      "emotion_detected": "frustrated",
      "emotion_confidence":0.85
    }
  ],
  "engagement_metrics": {
  "task_completion_rate": 90,
  "sustained_attention_seconds":120,
  "adaptation_events": 3,
   "successful_adaptations": 2

  }
}
```


**Fig. 4.** Example of Exercise Performance Log

To illustrate the practical implementation of the *Exercise Performance Logging and Evaluation* component, an example of session data recorded during the *Word Chain Game* is provided in JSON format as shown in Figure 4. This structured logging systematically captures critical metrics, including response accuracy, response time, speech recognition confidence, and emotional states detected via DeepFace. It further records task engagement metrics, such as task completion rate, sustained attention duration, and the effectiveness of real-time adaptive interventions. Additionally, incorporating complete verbal transcription of user responses would enable further linguistic analyses (e.g., fluency, complexity, word-finding difficulties, vocabulary diversity), thereby enriching the

understanding of cognitive performance. Such enhanced metrics could support the dynamic adaptation of robot behaviours, tailoring interactions based on real-time emotional and cognitive feedback. The detailed session data presented thus is an illustrative example of how performance metrics can guide real-time adaptive interventions while providing clinicians with comprehensive insights for personalized cognitive care.

Additionally, facial emotion recognition integrated into the Pepper robot was evaluated using real-time camera feed analyses. Tests confirmed high accuracy under optimal conditions, exemplified by successfully recognizing the participant's emotional expressions, such as detecting "happiness" with a confidence level of approximately 98%, as seen in Figure 5. This feature significantly enhanced the robot's adaptive and interactive capabilities, enabling personalized interventions based on real-time emotional feedback. However, it is essential to acknowledge that the testing conducted was preliminary and involved only the authors. Hence, findings regarding usability and effectiveness require cautious interpretation and further validation with diverse participant groups.

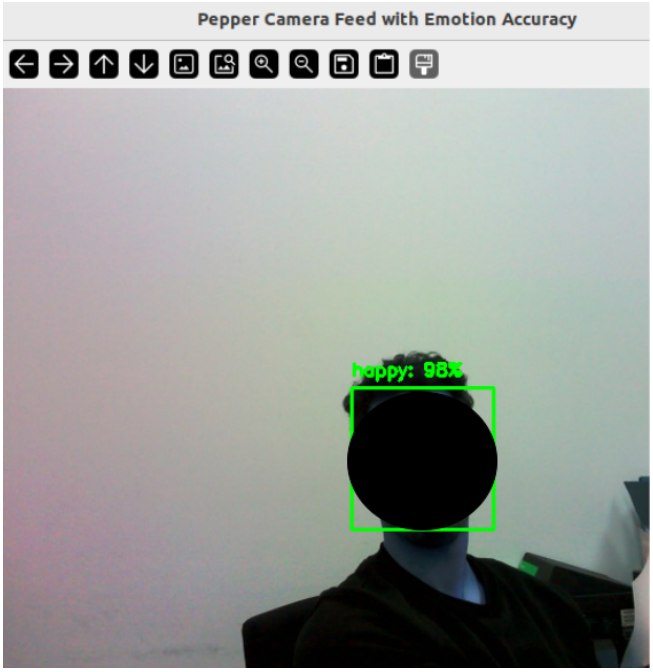


**Fig. 5.** Pepper's camera view with detected emotion – Happy

## 5 Limitations and Future Work

The evaluation presented in this study was preliminary, and although intended to test the technical feasibility and proof of concept of the system, was conducted exclusively by the authors, which limits the generalizability of the findings. Currently, the system incorporates only two cognitive exercises, the Word Association and Word Chain games, which constrains its applicability across broader cognitive domains. Additionally, the emotion recognition component, crucial for adaptive interactions, exhibited limitations in challenging conditions, such as low lighting or non-frontal face orientations. These limitations may affect the robustness and accuracy of emotional and cognitive assessments.

To comprehensively validate the MICA framework, in the future, we plan to carry out extensive evaluations with diverse participant groups, including older adults, individuals with cognitive impairments, and younger adults, while also considering gender and ethnicity. We will expand the cognitive test battery and repertoire, to ensure measurement of a broader range of targeted cognitive functions such as episodic and semantic memory, language, attention, and executive functions, and establishing correlations between the framework's cognitive and behavioural metrics and established neuropsychological assessments is essential for clinical acceptance. Collaborative approaches such as co-creation workshops involving patients, healthcare professionals, caregivers, and researchers can further refine the framework's clinical validity and usability. Enhancing the variety and complexity of cognitive exercises will improve personalization and user engagement. Finally, refining the multimodal perception component through advanced algorithms and integrating modalities like eye-tracking could enhance system performance.

## 6 Conclusion

This study has presented and demonstrated the novel MICA (Multimodal Intelligent Cognitive Assessment) framework integrated into the Pepper social robot, effectively leveraging multimodal emotional feedback and generative AI technologies to offer adaptive cognitive interventions. The cognitive exercises, including the Word Association and Word Chain games, illustrated the framework's promising potential for personalized cognitive training. The MICA framework represents an innovative approach to enhancing cognitive assessment and improving patient engagement. However, further research and development remain essential to overcome existing technical limitations, particularly in refining emotion recognition accuracy, expanding the diversity of cognitive exercises, and validating the framework through extensive testing with diverse participant groups.

**Acknowledgments.** This work is funded by the Strathclyde Studentship project "Intelligent Human-Robot Collaboration for Future Advanced Healthcare Applications" (2023-2026), under studentship number 3090.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. World Health Organization. Ageing and health - fact sheet. `https://www.who.int/news-room/fact-sheets/detail/ageing-and-health`, October 2024.
2. Shengyu Gu. A comparative study of increasing demand for health care for older people in china and the united kingdom. *World Sci. Res. J*, 6:218–251, 2020.
3. World Health Organization. Dementia. `https://www.who.int/news-room/fact-sheets/detail/dementia`, Mar. 31 2025. Accessed: 4-Apr-2025.
4. Shuyang Yu and Meng Wei. The influences of community-enriched environment on the cognitive trajectories of elderly people. *International Journal of Environmental Research and Public Health*, 18(16):8866, 2021.
5. Jaime Andres Rincon Arango, Cedric Marco-Detchart, and Vicente Javier Julian Inglada. Personalized cognitive support via social robots. *Sensors (Basel, Switzerland)*, 25(3):888, 2025.
6. Róbert Sabo, Štefan Beňuš, Viktória Kevická, Marian Trnka, Milan Rusko, Sakhia Darjaa, and Jay Kejriwal. Towards the use of social robot furhat and generative ai in testing cognitive abilities. *Human Affairs*, 34(2):224–243, 2024.
7. Mohamed Adlan Ait Ameur, Erfu Yang, and Scott Brady. Advancing healthcare through intelligent human-robot collaboration: Overview, challenges and opportunities. In *2024 29th International Conference on Automation and Computing (ICAC)*, pages 1–6, 2024.
8. Alistair A Vogan, Fady Alnajjar, Munkhjargal Gochoo, and Sumayya Khalid. Robots, ai, and cognitive training in an era of mass age-related cognitive decline: a systematic review. *Ieee Access*, 8:18284–18304, 2020.
9. Miguel A Salichs, Álvaro Castro-González, Esther Salichs, Enrique Fernández-Rodicio, Marcos Maroto-Gómez, Juan José Gamboa-Montero, Sara Marques-Villarroya, José Carlos Castillo, Fernando Alonso-Martín, and Maria Malfaz. Mini: a new social robot for the elderly. *International Journal of Social Robotics*, 12:1231–1249, 2020.
10. Konstantinos Tsiakas, Maher Abujelala, and Fillia Makedon. Task engagement as personalization feedback for socially-assistive robots and cognitive training. *Technologies*, 6(2):49, 2018.
11. Kari Blindheim, Mads Solberg, Ibrahim A Hameed, and Rigmor Einang Alnes. Promoting activity in long-term care facilities with the social robot pepper: a pilot study. *Informatics for Health and Social Care*, 48(2):181–195, 2023.
12. Katie Winkle, Praminda Caleb-Solly, Ailie Turton, and Paul Bremner. Social robots for engagement in rehabilitative therapies: Design implications from a study with therapists. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction*, pages 289–297, 2018.
13. Chai M Tyng, Hafeez U Amin, Mohamad NM Saad, and Aamir S Malik. The influences of emotion on learning and memory. *Frontiers in psychology*, 8:235933, 2017.
14. Sefik Ilkin Serengil and Alper Ozpinar. Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–4. IEEE, 2021.

15. Alaa Abd-Alrazaq, Dari Alhuwail, Eiman Al-Jafar, Arfan Ahmed, Farag Shuweihdi, Shuja Mohd Reagu, and Mowafa Househ. The effectiveness of serious games in improving memory among older adults with cognitive impairment: systematic review and meta-analysis. *JMIR serious games*, 10(3):e35202, 2022.
16. Pantelis Pergantis, Victoria Bamicha, Charalampos Skianis, and Athanasios Drigas. AI chatbots and cognitive control: Enhancing executive functions through chatbot interactions: A systematic review. *Brain Sciences*, 15(1):47, 2025.
17. Jim Lumsden, Elizabeth A Edwards, Natalia S Lawrence, David Coyle, Marcus R Munafò, et al. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR serious games*, 4(2):e5888, 2016.
18. Giovanna Castellano, Berardina De Carolis, Nicola Macchiarulo, and Olimpia Pino. Detecting emotions during cognitive stimulation training with the pepper robot. In *Human-Friendly Robotics 2021: HFR: 14th International Workshop on Human-Friendly Robotics*, pages 61–75. Springer, 2022.
19. Ka Ki Chow, Chun Sing Ip, Chun To Yau, Jingqiang Zeng, and Junpei Zhong. A systematic review of using human-robot interaction for cognitive training for elderly with mild cognitive impairment. In *International Conference on Intelligent Robotics and Applications*, pages 30–41. Springer, 2025.
20. Antonio Blanco Castillo, Pedro Núñez, Jackie Lee Chong Ojeda, and Alicia Cordón. Enhancing cognitive therapies for older adults with ebo: An autonomous social robot system integrating llms and therapist-in-the-loop. In *Proceedings of the XXIV Workshop of Physical Agents: September 5-6, 2024*, pages 31–45. Universidad de Alicante/Universitat d'Alacant, 2024.
21. Antonio Andriella, Carme Torras, and Guillem Alenya. Cognitive system framework for brain-training exercise based on human-robot interaction. *Cognitive Computation*, 12(4):793–810, 2020.
22. Zilong Wang, Nan Chen, Luna K Qiu, Ling Yue, Geli Guo, Yang Ou, Shiqi Jiang, Yuqing Yang, and Lili Qiu. The potential and value of ai chatbot in personalized cognitive training. *arXiv preprint arXiv:2410.19733*, 2024.
23. Hang Su, Wen Qi, Jiahao Chen, Chenguang Yang, Juan Sandoval, and Med Amine Laribi. Recent advancements in multimodal human–robot interaction. *Frontiers in Neurorobotics*, 17:1084000, 2023.
24. Uday Kamath, Kevin Keenan, Garrett Somers, and Sarah Sorenson. Prompt-based learning. In *Large Language Models: A Deep Dive: Bridging Theory and Practice*, pages 83–133. Springer, 2024.
25. Ggaliwango Marvin, Nakayiza Hellen, Daudi Jjingo, and Joyce Nakatumba-Nabende. Prompt engineering in large language models. In *International conference on data intelligence and cognitive informatics*, pages 387–402. Springer, 2023.
26. Alice Kwak, Clayton Morrison, Derek Bambauer, and Mihai Surdeanu. Classify first, and then extract: Prompt chaining technique for information extraction. In *Proceedings of the Natural Legal Language Processing Workshop 2024*, pages 303–317, 2024.
27. Google Cloud. Speech-to-text documentation. `https://cloud.google.com/speech-to-text`, 2024. Accessed: 2025-06-01.
28. Erik Billing, Julia Rosén, and Maurice Lamb. Language models for human-robot interaction. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 905–906, 2023.