Dissertation

Effects of Cognitive Trainings with the Socially Assistive Robot “Pepper” vs. Tablet in Depressed Psychiatric Inpatients: A Randomized Controlled Pilot Trial

submitted by

Alfred Alois Häussl BScN. MSc

for the Academic Degree of

Doctor of Medical Science

(Dr. scient. med.)

at the

Medical University of Graz

Department of Psychiatry and Psychotherapeutic Medicine

under the supervision of

Univ.-Prof. Priv.-Doz. Dr. med. univ. et scient. med. Eva REININGHAUS, MBA

2026

Declaration

““I hereby confirm that the present diploma thesis is the result of my own independent scholarly work. I also confirm that in all cases, where material from the work of others (in books, articles, essays, dissertations, and on the internet) is acknowledged, quotations and paraphrases are clearly indicated. No material other than that cited in the reference list has been used. I have read and understood the Medical University's regulations and procedures concerning plagiarism. Furthermore, I hereby declare that if artificial intelligence (AI) tools were used for the generation and/or correction of certain text passages in the creation of this work, such employment was conducted in compliance with ethical principles, academic integrity, and the regulations of my university. Additionally, it was ensured that this usage was transparently disclosed and appropriately attributed."

The following generative AI tools were used to optimize the language of the text:

- ChatGPT 5.2, OpenAI, 02/102/2026; https://www.chatgpt.com
- Grammarly Pro, Grammerly, Inc., 02/02/2026; https://www.grammarly.com
- Paperpal Prime Plan, Editage, 12/12/2025; https://www.paperpal.com

Graz, February 04th 2026 Alfred Häussl BScN MSc m.p.

Disclosures

This thesis was written at the doctoral school of “Lifestyle-related diseases”.

Parts of this thesis have been published in the following article:
Haeussl A[1,2], Zwigl I[1], Stojec L[1], Smolak I[1], Stijic M[3], Stross T[1], Lenger M[1], Fellendorf FT[1], Guggemos S[1], Schoenthaler EMD[1], Pszeida M[4], Orgel T[4], Draxler S[4], Schneeberger M[4], Russegger S[4], Zuschnegg J[5], Steindl D[5], Schultz A[5], Schuessler S[5], Macher M[6,7], Paletta L[4], Dalkner N[1] and Reininghaus EZ[1] (2026) Evaluating usability and acceptance of a socially assistive robot supported cognitive training for depression – results of the randomized controlled pilot study ‘AMIGA’. *Front. Psychiatry* 16:1661730. doi: 10.3389/fpsyt.2025.1661730.

Author affiliations

[1] Department of Psychiatry, Psychosomatic and Psychotherapy, Clinical Division of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Graz, Austria

[2] Department of Healthcare and Nursing, Department of Health Sciences, Faculty of Health and Social Sciences, USTP – University of Applied Sciences St. Poelten, Sankt Poelten, Austria

[3] Department of Neurology, Division of Neurogeriatrics, Medical University of Graz, Graz, Austria

[4] DIGITAL – Institute for Digital Technologies, Joanneum Research Forschungsgesellschaft mbH, Graz, Austria

[5] Institute of Social Medicine and Epidemiology, Medical University of Graz, Graz, Austria

[6] Department of Media and Digital Technologies, Faculty of Engineering and Business, USTP – University of Applied Sciences St. Poelten, Sankt Poelten, Austria

[7] Humanizing Technologies GmbH, Vienna, Austria

All co-authors agree to the inclusion of their published data in the dissertation. Written statements are submitted together with the thesis.

Permission to reproduce tables/figures form own (or third-party) publications has been granted and stated.

Acknowledgements

Doctoral student Alfred Häussl BScN MSc received funding from the “Austrian FFG femtech programme co-financed by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK)” under the grant No. FO999892829 and the Medical University of Graz through the Doctoral School “Lifestyle-related Diseases”.

It is with great pleasure and gratitude that I would like to take this opportunity to thank you for your valuable support during my doctoral thesis. This project would not have taken the same successful course without the help and encouragement of some special people.

A big thank you goes to my three supervisors, Univ.-Prof. Priv.-Doz. Dr. Eva Reininghaus, MD MBA, Research Prof. Priv.-Doz. Mag. Dr. Nina Dalkner and DI Dr. Lucas Paletta, who not only impressed me with their professional expertise, but were also always available to answer my questions and concerns. Their guidance and constructive advice contributed significantly to my dissertation reaching the desired standard.

I would also like to thank my family, especially my parents, who have always believed in me. Their unconditional support and words of encouragement have always given me the motivation I needed to persevere, even during challenging phases. I would also like to express my sincere gratitude to my brother Peter and my grandparents for their encouraging and supportive words.

Finally, I would like to thank my girlfriend Ina from the bottom of my heart. Her energetic support, both on a professional and personal level, has been invaluable to me. Her patience, encouragement and understanding helped me to tackle the challenges of the dissertation with confidence.

Table of content

Abbreviations and their explanations

ACC = Accessibility
ADHD = Attention-Deficit-Hyperactivity-Disorder
ADLs = Activities of Daily Living
AI = Artificial Intelligence
ANCOVA = Analysis of Covariance
ANOVA = Analysis of Variance
ANX = Anxiety
AP-F1 = Subscale Prospective Memory Performance
AP-F2 = Subscale Maintaining Focused Attention
API = Application Programming Interface
APS-20 = Overall Score of Attention and Performance Self-Assessment
APSA = Attention and Performance Self-Assessment
ASD = Autism-Spectrum-Disorder
BDI-II = Beck Depression Inventory Revision
BSI-18 = Brief Symptom Inventory 18
CBT = Cognitive-Behavioral Therapy
CHAL = Challenge
CI = Confidence Interval
CMS = Content Management System
CON = Confidence
Corr. = Corrected
CUR = Curiosity
df = Degrees of Freedom
DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
EOU = Ease of Use
F = F-statistics
FoF = Fear of Failure
GSQ = Godspeed Questionnaire
HADS = Hospital Anxiety and Depression Scale
HAM-D = Hamilton Depression Rating Scale
HRI = Human-Robot Interaction
ICC = Intraclass Correlation

ICD-10 = International Statistical Classification of Diseases and Related Health Problems Version 10

IDAS-II = Inventory of Depression and Anxiety Symptoms – German Version

IMM = Immersion

INT = Interest

IQ = Intelligence Quotient

ITU = Intention to Use

M = Mean

MADRS = Montgomery–Åsberg Depression Rating Scale

MANCOVA = Multivariate Analysis of Covariance

MANOVA = Multivariate Analysis of Variance

N = Total Sample Size

n = Sample size of a specific group within the total sample size

p = Significance Level

partial n^2 or n^2_p= Partial Eta Squared

PoS = Probability of Success

PRI = Pride

PU = Perceived Usefulness

PUE = Perceived Ease of Use

QAM = Questionnaire on Achievement Motivation

QCM = Questionnaire on Current Motivation

QoL = Quality of Life

RCT = Randomized Controlled Trial

SAR = Socially Assistive Robot

SARs = Socially Assistive Robots

SD = Standard Diviation

SHA = Shame

Sig = Significane

SKE = Skepticism

SQL = Structured Query Language

Sub = Subscale

SUS = System Usability Scale

T = T-statistics for independent sample t-test

t_0 = Time of Measurement 1 (Baseline)

t_1 = Time of Measurement 2 (after intervention)

TAM = Technology Acceptance Model

ToM = Time of Measurement

TRA = Theory of Reasoned Action

TUI = Technology Usage Inventory

USE = Usefulness

VR = Virtual Reality

WHO = World Health Organization

XR = Extended Reality

List of figures

List of tables

Abstract in German

Hintergrund: Depression zählt zu den häufigsten psychiatrischen Erkrankungen weltweit und ist mit erheblichen Einschränkungen der Lebensqualität sowie einer hohen Belastung für Gesundheitssysteme verbunden. Neben etablierten Therapieformen gewinnen digitale Interventionen wie kognitives Training mit sozial assistierenden Robotern (SARs) zunehmend an Bedeutung.

Ziel: Ziel dieser Dissertation war es, die Effekte eines kognitiven Trainings mit dem SAR „Pepper“ im Vergleich zu einem Tablet-basierten Training auf depressive Symptome, Aufmerksamkeit und Leistung, Akzeptanz und Benutzerfreundlichkeit sowie Motivation bei stationär behandelten Patient*innen mit Depression zu untersuchen. Zusätzlich wurden geschlechts- und altersbezogene Unterschiede analysiert.

Methode: Es wurde eine randomisierte kontrollierte Pilotstudie mit 32 stationären Patient*innen mit Depression durchgeführt. Die Teilnehmenden absolvierten zwei Trainingseinheiten entweder mit dem SAR „Pepper“ mit motivierendem Feedback oder mit einem Tablet ohne Feedback. Die Erhebung erfolgte mittels validierter und selbsterstellter Fragebögen.

Ergebnisse: Beide Gruppen zeigten nach dem Trainingszeitraum eine Reduktion depressiver Symptome ohne signifikante Unterschiede. Auch für Aufmerksamkeit und Leistung ergaben sich keine gruppenspezifischen Effekte. Die SAR-Gruppe erzielte jedoch signifikant höhere Bewertungen in den Skalen „Benutzerfreundlichkeit“ und „wahrgenommenen Einfachheit“. Geschlechtsspezifische Unterschiede zeigten sich unabhängig von der Interventionsform insbesondere in der Selbstwahrnehmung von Aufmerksamkeit und Leistung sowie in motivationalen Dimensionen. Qualitative Rückmeldungen bestätigten die hohe Akzeptanz und die als unterstützend erlebte soziale Interaktion mit dem Roboter.

Schlussfolgerung: Die Ergebnisse belegen die Machbarkeit und hohe Benutzerfreundlichkeit von SAR-gestütztem kognitivem Training im akuten psychiatrischen Setting. Für Aussagen zu klinischen Effekten sind längere Interventionszeiträume und größere Stichproben erforderlich.

Schlüsselwörter: Sozial assistierende Roboter (SAR), Depression, Technologieakzeptanz, Benutzerfreundlichkeit, Kognitives Training, Aufmerksamkeit und Leistung, Motivation, Psychiatrie, Klinisches Setting, Geschlechterunterschiede, Altersunterschiede

Abstract in English

Background: Depression is one of the most prevalent psychiatric disorders worldwide and is associated with substantial impairments in quality of life as well as a high burden on health care systems. In addition to established treatment approaches, digital interventions such as cognitive training supported by socially assistive robots (SARs) are gaining increasing relevance

Aim: The aim of this doctoral dissertation was to investigate the effects of cognitive training delivered by the socially assistive robot “Pepper” compared to tablet-based training on depressive symptoms, attention and performance, acceptance and usability, and motivation in psychiatric inpatients with depression. In addition, sex- and age-related differences were examined.

Methods: A randomized controlled pilot study was conducted with 32 psychiatric inpatients diagnosed with depression. Participants completed two training sessions either with the SAR “Pepper” providing motivational feedback or with a tablet without feedback. Outcomes were assessed using validated and self-developed questionnaires.

Results: Both groups showed a reduction in depressive symptoms following the training period, with no significant group differences. No group-specific effects were observed for attention and performance. However, the SAR group reported significantly higher scores in “usability” and “perceived ease of use” scales. Sex-specific differences emerged irrespective of the intervention, particularly in self-reported attention and performance as well as in motivational dimensions. Qualitative feedback confirmed high acceptance and the socially supportive interaction experienced with the robot.

Conclusion: The findings demonstrate the feasibility and high usability of SAR-supported cognitive training in acute psychiatric inpatient settings. To draw conclusions regarding clinical effects, longer intervention periods and larger sample sizes are required.

Keywords: Socially assistive robots (SAR), depression, technology acceptance, usability, cognitive training, attention and performance, motivation, psychiatry, clinical setting, sex differences, age differences

1. Introduction

1.1 Background

Depression is a mental health disorder that significantly impacts individuals, families, and healthcare systems worldwide (1). According to the World Health Organization (WHO), depression affects more than 280 million people globally, making it a leading cause of disability and a major contributor to the global burden of disease (1). In Austria, approximately 9.8% of the population experience depressive episodes annually, with women (11.5%) being more affected than men (7.9%) (2). The COVID-19 pandemic has further exacerbated the prevalence of depression, particularly among vulnerable groups, such as individuals with chronic illnesses (3,4).

Depression, as defined by the International Statistical Classification of Diseases and Related Health Problems Version 10 (ICD-10), is characterized by pervasive changes in mood, such as diminished interest or pleasure, reduced energy, fatigue, and recurrent thoughts of death or suicide. Cognitive, emotional, and physical impairments often accompany these symptoms, significantly affecting daily functioning (5). Beyond its clinical symptoms, depression adversely affects quality of life and psychosocial functioning, commonly contributing to social withdrawal and lower productivity. Summarized, it is a severe multifactorial disorder influenced by genetic, environmental, and psychological factors (6).

Treatment typically involves a combination of pharmacological interventions, such as antidepressants, and psychotherapeutic approaches, psychotherapy, cognitive-behavioral therapy (CBT), social support, and lifestyle modifications (6–8). However, a central challenge in managing depression is overcoming the lack of motivation that often hinders therapy adherence and lifestyle modifications. This can exacerbate depressive symptoms, perpetuating a cycle of cognitive and emotional decline (9).

Cognitive training has emerged as a promising approach to address depression challenges such as low attention and memory function often associated with neglect of essential daily activities, including self-care, family responsibilities, and professional obligations. There are different types of cognitive trainings, but in total they comprise structured exercises aimed at improving cognitive and emotional regulation (10).

Most cognitive training programs are designed to improve memory, attention, emotional perception, and executive function, which are often impaired in individuals with depression (10–13). These interventions aim to target neurocognitive deficits that contribute to the persistence

of depressive symptoms and functional impairment. By enhancing cognitive capacities, such programs can help individuals better manage daily tasks, improve emotional resilience, and reduce relapse rates (14,15). Moreover, integrating cognitive training with standard treatments may offer synergistic benefits, leading to more comprehensive and sustained recovery outcomes (9,16–20). Research has shown that cognitive training, particularly in computer-based formats, can significantly reduce depressive symptoms and improve motivation (10,21–23).

Modern cognitive training programs like “Lumosity”, “CogniFit”, “Peak”, and “Cogmed” (24) incorporate adaptive features and real-time feedback, which enhance user engagement and effectiveness by tailoring tasks to the individual's performance level (25). These dynamic adjustments help maintain an optimal challenge, preventing frustration or boredom, and promote sustained motivation throughout the training process. Additionally, the integration of gamified elements and interactive interfaces further supports adherence and facilitates the transfer of cognitive improvements to everyday functioning (26).

While cognitive training has proven effective in addressing depressive symptoms and improving cognitive resilience, the integration of advanced digital technologies holds significant promise to further revolutionize mental health care (10,21–23). The ongoing digitalization of healthcare has opened new pathways for delivering personalized, scalable, and engaging interventions that can be tailored to individual patient needs. Among these innovations, technologies such as socially assistive robots (SARs) are playing a pivotal role by providing interactive and empathetic support, which enhances the therapeutic experience (26). These technological advancements not only improve the delivery of therapeutic interventions but also help overcome critical challenges related to accessibility, adherence, and sustained patient engagement, particularly in underserved or remote populations (27).

The following chapter delves deeper into the broader context of digitalization, examining how emerging technologies, particularly SARs, are fundamentally transforming therapeutic and rehabilitative practices (28). It will provide a comprehensive overview of the unique features of SARs, emphasizing their capacity to deliver personalized, multisensory experiences that adapt to individual patient needs and preferences. Furthermore, the chapter will explore the growing integration of SARs within traditional psychiatric treatment frameworks, highlighting their potential to enhance patient engagement, emotional support, and treatment adherence (29,30). By investigating both technological advancements and clinical applications, this section aims to elucidate the evolving role of SARs in mental health care and rehabilitation.

1.2 Digitalization

Digital technologies such as robots, virtual reality, and extended reality are becoming increasingly omnipresent in our modern world, fundamentally transforming various aspects of daily life, work, and social interaction. Their pervasive integration is reshaping how individuals communicate, learn, and access services, thereby influencing societal norms and behaviors on a broad scale. This rapid technological advancement presents a dual-edged scenario, offering significant opportunities for innovation and improvement in numerous fields, including healthcare, while simultaneously introducing new risks and ethical considerations that require careful evaluation and management (26).

Within the healthcare system, the potential applications of these digital technologies are vast, ranging from enhanced diagnostic tools and personalized treatment plans to immersive rehabilitation programs and remote patient monitoring (31). However, the adoption of such innovations must be approached with a critical and comprehensive perspective, ensuring that technological benefits are maximized without compromising patient safety, privacy, or equity of access. Active, in-depth discussions involving healthcare professionals, policymakers, technologists, and patients are essential to navigate the complexities associated with integrating these tools into clinical practice effectively and ethically (32).

Moreover, the evolving landscape of digital health demands ongoing research and evaluation to understand the long-term impacts of these technologies on patient outcomes and healthcare delivery models. This includes assessing not only clinical effectiveness but also factors such as user engagement, accessibility for diverse populations, and the potential to reduce disparities in healthcare access. By fostering a balanced dialogue that weighs both the promising advantages and inherent challenges, stakeholders can collaboratively shape a future where digital technologies contribute meaningfully and responsibly to improved health and well-being (26).

1.2.1 Social Assistive Robots (SAR)

Socially assistive robotics, an emerging and rapidly evolving field of technology, rigorously examines both the psychological underpinnings and the technical architectures required for meaningful human-robot interaction (HRI) (33–35). Unlike industrial, SARs are intentionally designed to provide social, emotional, and practical functional support, positioning them as uniquely valuable assets, particularly within healthcare and specialized caregiving environments.

The development of SARs is rooted in understanding social cues, emotional intelligence simulation, and trust-building mechanisms (33–35).

Their applications have demonstrated particularly promising results in the domain of elderly care. Here, SARs move beyond simple task automation to actively foster social interaction, thereby mitigating the pervasive issue of chronic loneliness among older adults. They achieve this by engaging users in structured activities, providing conversational prompts, or simply offering a consistent, non-judgmental presence. Furthermore, SARs assist in activities of daily living (ADLs) through gentle reminders for medication adherence, monitoring vital signs, or guiding users through simple physical exercises (36). The technical challenge here lies in ensuring the robot's responses are contextually appropriate and personalized, moving beyond scripted interactions to achieve genuine perceived social value.

To achieve robust social acceptance, SARs must possess advanced social communication skills (37). This encompasses more than just verbal fluency; it requires nuanced abilities such as recognizing and responding to subtle human emotional cues (e.g., detecting frustration through vocal tone or facial expressions), and employing context-aware dialogue (38,39). Furthermore, SARs must consistently exhibit trustworthiness. This is often built through predictable, reliable performance and transparency regarding their capabilities and limitations. For instance, a SAR designed to remind an elderly user about medication must execute this task flawlessly every time to maintain that trust (40). Equally critical is maintaining non-invasive behavior. This relates to respecting personal space boundaries, a concept that varies culturally but is vital for comfort. If a SAR encroaches too closely during a task, acceptance plummets, regardless of its utility (41).

Physical size of SARs plays a critical role in their acceptance by users (42). Smaller, less imposing designs, often resembling familiar objects or having soft, non-threatening aesthetics, generally facilitate easier integration into domestic or clinical settings compared to large, anthropomorphic machines. While initial interactions with SARs may feel unfamiliar or even awkward – a phenomenon sometimes termed the "uncanny valley" effect (43) – studies show that users quickly develop natural communication patterns and trust in these robots, leading to increased acceptance (44).

SARs have been effectively utilized to enhance the quality of life (QoL) and independence of various populations. For example, they serve as cognitive aids and companions for individuals with dementia (45). In this context, SARs like memory aids can reduce anxiety associated with memory loss by providing timely prompts for daily activities or offering comforting, repetitive interaction that mitigates feelings of isolation. More recently, advanced SARs are being

deployed to assist in rehabilitation following stroke, offering repetitive, encouraging exercise guidance that human therapists might not be available to provide continuously (46). This consistent, tireless support directly contributes to measurable gains in patient independence.

Unlike virtual interventions limited strictly to screens, a SAR is physically present in the room, providing a genuinely multisensory interaction experience. SARs can engage users through multiple modalities, including auditory feedback (like spoken encouragement or calming sounds), salient visual cues (such as gestures or displaying relevant information on a screen integrated into the robot), and tactile interaction (such as a gentle pat or holding a user's hand, as demonstrated in some eldercare prototypes). This versatility significantly enhances its suitability for a broad range of therapeutic and educational settings, accommodating users with varying physical abilities or cognitive loads (26).

Furthermore, the capacity of a SAR to socially engage with individuals allows it to present highly personalized interventions, such as structured skills training (e.g., practicing social cues for individuals with Autism Spectrum Disorder) or continuous health tracking (e.g., monitoring medication adherence or activity levels). This is similar to mobile apps but delivered in a far more interactive, tangible, and engaging format that leverages social presence (29,47). The social capabilities of a SAR can be specifically tailored and enhanced through sophisticated AI algorithms to support individuals in achieving complex mental and behavioral health goals, such as reducing social isolation or improving emotional regulation skills (48,49). A key differentiator is the SAR's ability to build personalized affective relationships with its users by for example remembering past conversations or displaying consistent empathetic responses and thereby supporting them more effectively in pursuing their treatment objectives. By doing so, a SAR actively contributes to the establishment of an emotionally supportive environment, which can play a critical, measurable role in improving patient adherence and clinical outcomes (30).

In an era where healthcare resources are increasingly scarce and access remains uneven, SARs hold significant potential to complement traditional psychiatric treatments. These robots offer an innovative, scalable approach to addressing critical gaps in mental health care by providing consistent, immediate supplementary support to both patients needing ongoing monitoring and overburdened healthcare providers (50). For example, a SAR can handle routine check-ins or deliver standardized CBT exercises, freeing up human therapists for complex clinical work. By integrating advanced sensing technology with thoughtful human-centered design principles a SAR exemplifies how robotics can positively impact patient experiences and improve overall care delivery efficiency. As SARs continue to evolve with

improved natural language processing and emotional recognition capabilities, their inherent ability to adapt dynamically to users' evolving needs and preferences ensures their growing relevance and integration into routine therapeutic and rehabilitative contexts (26).

1.3 Theoretical Concept of Acceptance

This chapter explores the theoretical underpinnings of acceptance, its specific relevance to SARs and tablets in clinical contexts, and its application within the framework of the Technology Acceptance Model (TAM).

Acceptance is a crucial concept in understanding the adoption and use of new technologies across various domains, particularly in healthcare. The term "acceptance" refers to the willingness of individuals to engage with and utilize a technology, reflecting their perceptions of its utility, ease of use, and overall value (51). In healthcare, the acceptance of innovative tools is paramount, as it directly impacts the effectiveness of interventions, adherence to treatments, and patient outcomes.

Acceptance in this thesis means when …

(1) there is a positive attitude towards the innovation.
(2) there is a behavioral intention to use the innovation.
(3) the innovation is used. (52)

Of course, many different factors play a role here, for example, the characteristics of the innovation and the potential benefit for the user are very important. The greater the benefit for the users and the more open-minded and educated the users are, the higher the acceptance is. The way in which the innovation is offered, i.e., the earlier the users are informed and involved, and the more opportunities the users have for co-determination, the higher the acceptance is (52).

The Technology Acceptance Model, which is the key concept of this study, is described in more detail.

1.3.1 Technology Acceptance Model

The Technology Acceptance Model (TAM) is a foundational and highly influential multidimensional construct within information systems research, designed to capture the complex interplay of cognitive, emotional, and behavioral aspects influencing an individual's decision to adopt new technology. Developed in 1986 by Fred Davis, the TAM quickly established itself as a cornerstone of technology acceptance research (51,53). This model's genesis lies in the earlier *"Theory of Reasoned Action (TRA)"* formulated by Fishbein and Ajzen in 1975 (54). The TRA posits that an individual's behavior is primarily determined by their behavioral intention, which is, in turn, shaped by their attitudes toward the behavior and subjective norms (54,55).

Davis adapted the TRA to the specific context of information technology adoption, streamlining the framework to focus on two primary determinants that directly influence a user's intention to use a system, thereby predicting actual usage behavior: *"Perceived Usefulness (PU)"* and *"Perceived Ease of Use (PUE)"* (Davis, 1986).

"Perceived Usefulness (PU)" is operationally defined as the degree to which an individual believes that using a particular system would enhance or improve their job performance or task accomplishment (51). For instance, if an employee believes that a new system will allow them to process customer orders 30% faster this constitutes high Perceived Usefulness. A high degree of PU creates a strong positive motivation, directly leading to a higher intention to use the technology, as users see a clear instrumental benefit.

"Perceived Ease of Use (PUE)", conversely, refers to the extent to which a person anticipates that the effort required to use the technology will be minimal or manageable (51). This factor addresses the cognitive load and user experience. For example, an interface with intuitive navigation and clear labeling contributes significantly to high PUE. Crucially, PUE influences PU; if a system is perceived as too difficult to learn or operate (low PUE), users are less likely to believe it can actually improve their performance (low PU), even if the underlying capabilities exist. Therefore, PUE acts as an antecedent to PU, as well as having a direct influence on the intention to use the technology. Together, these two factors form the core predictive power of the original TAM framework (51). The TAM is shown in **Figure 1**.

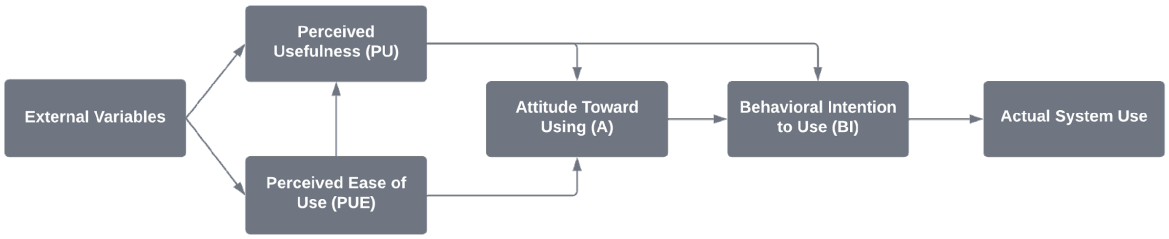

Figure 1: *Technology Acceptance Model (TAM) by Davis (1989) created by Alfred Haeussl*

1.3.2 Acceptance in Healthcare Context

In healthcare, the acceptance of technology is influenced by a confluence of unique considerations, which extend beyond mere usability. These factors critically include the inherent “patient vulnerability” (where reliance on technology can carry significant risks if errors occur), “ethical implications” surrounding autonomy, consent, and algorithmic bias, and the profound potential for technological interventions to “disrupt traditional care paradigms” (26). For instance, the introduction of AI-driven diagnostic tools necessitates careful navigation of accountability when interpretations differ from human judgment.

Crucially, studies consistently demonstrate that technologies exhibiting high acceptance rates are significantly more likely to be integrated successfully into complex clinical workflows and, consequently, result in positive patient outcomes (56–60). High acceptance, often measured through metrics like Technology Acceptance Model (TAM) or System Usability Scale (SUS) scores or adoption frequency, directly correlates with sustained benefit realization. Conversely, low acceptance leads to underutilization or outright abandonment of potentially transformative technology, regardless of its established clinical efficacy or cost-saving potential (61).

The healthcare sector presents distinct, layered challenges to technology acceptance. “Patient reluctance” often stems from several interconnected issues, including anxiety related to interacting with unfamiliar digital interfaces (technophobia), a genuine lack of technical proficiency, particularly among older demographics, low technology-related self-efficacy, and cognitive overload when digital systems are perceived as complex or unintuitive. These factors are frequently reinforced by negative past experiences with digital health technologies, fears of losing control over personal health decisions, and social influences such as skeptical attitudes within family or peer networks. In addition, deeply rooted concerns about privacy and the security of highly sensitive protected health information remain a major barrier to acceptance (26). For example, patients may hesitate to use a patient portal if they fear unauthorized access to their medical records.

Similarly, healthcare providers such as physicians, nurses, and allied health professionals may actively resist adopting new digital tools if they perceive them as time-consuming, cognitively demanding, or adding unnecessary “clicks” to already burdened workflows. Resistance is further amplified when technologies are experienced as complex to master, misaligned with established patient care goals, or when prior negative encounters with poorly designed systems have reduced trust and confidence in new solutions (62). Concerns about loss of professional autonomy and social influences within clinical teams can additionally shape attitudes toward adoption (62). Overcoming these multifaceted barriers therefore requires a nuanced, evidence-based understanding of the psychological, operational, and systemic factors that drive technology acceptance across diverse user groups, necessitating tailored strategies that address the distinct needs and priorities of patients, clinicians, and healthcare administrators alike.

1.3.3 Socially assistive Robots and Acceptance

SARs represent a novel and rapidly evolving category of technology meticulously designed to provide crucial social and emotional support while simultaneously assisting users in achieving specific, measurable goals, such as structured cognitive training modules or personalized physical rehabilitation protocols. Prominent examples of operational SARs, such as SoftBank Robotics' "Pepper" or the therapeutic robot "PARO" (a robotic seal used extensively in dementia care), effectively combine interactive communication features, deliberately anthropomorphic or zoomorphic design elements, and sophisticated artificial intelligence algorithms (often incorporating natural language processing (NLP) and machine learning) to create deeply engaging and highly personalized user experiences (28).

Crucially, the successful integration and acceptance of SARs are shaped by a unique constellation of factors that are distinct from those influencing the adoption of traditional, non-interactive technologies like standard tablets or desktop computers (26). The anthropomorphic qualities inherent in SAR design play a profoundly significant role in shaping user perceptions and willingness to interact. Research has demonstrated that the inclusion of human-like features, such as expressive eyes, facial mimicry, or recognizable body language, can substantially increase user trust, foster empathy, and deepen engagement, ultimately leading to demonstrably higher acceptance rates, especially among vulnerable populations like the elderly (50,63).

However, these very human-like features introduce a significant design paradox. They can also provoke feelings of unease, skepticism, or even aversion. This effect occurs when robots

appear almost, but not entirely natural, triggering negative emotional responses in human observers (64).

In sensitive clinical settings, the acceptance and usability of SARs are particularly relevant given their immense potential to enhance therapeutic outcomes beyond what traditional methods alone can achieve. Empirical studies have provided compelling evidence that SARs can effectively augment cognitive training programs by providing consistent, non-judgmental practice partners; improve social interaction skills, particularly for individuals on the autism spectrum; and significantly reduce reported feelings of loneliness and the associated social stigma often accompanying mental health challenges. For example, studies involving SARs in elder care have documented measurable improvements in mood states following consistent interaction sessions (65,66). However, their adoption is often limited by factors such as high costs, technical reliability, and the need for specialized training among healthcare providers (26).

1.3.4 Acceptance of Tablets in Healthcare

Unlike SARs, tablets are well-established technologies with widespread acceptance and proven integration within diverse healthcare settings. Their portability, characterized by lightweight designs and long battery life suitable for bedside or home use, combined with highly intuitive, user-friendly interfaces, make them exceptionally valuable tools for delivering a wide array of clinical interventions, including digital therapeutics and cognitive training programs. For instance, tablets are routinely used for administering standardized cognitive assessments, digitally, or for delivering structured physical therapy exercises via video instruction (61).

Tablets are often perceived as significantly less intimidating and more familiar than novel technologies like SARs, which directly contributes to higher acceptance rates among both elderly patients and busy clinical providers (62). This acceptance is rooted in technological familiarity; nearly all users have prior personal or professional experience with smartphones or similar touchscreen devices, effectively minimizing the initial learning curve and boosting confidence in the device's utility for health management (61). The widespread commercial availability and standardization of operating systems (like iOS or Android) further enhance this acceptance, as IT departments are already equipped to manage and secure these platforms. The familiarity of tablets, stemming from their pervasive presence in daily life, plays a crucial role in bridging the gap between consumer technology and clinical application (26,67).

However, despite these substantial advantages in accessibility and ease of use, tablets possess limitations when compared to immersive technologies. Specifically, standard tablets

often lack the advanced interactive, spatial computing, and high-fidelity motivational features characteristic of SARs. The absence of deep immersion and advanced feedback mechanisms may limit the tablet's sustained effectiveness in certain demanding applications, particularly those requiring high levels of engagement or behavioral modification over extended durations, such as adherence monitoring in long-term rehabilitation or complex therapy (26).

1.3.5 Comparative Acceptance of SAR and Tablets

The comparison between SARs and traditional tablets reveals crucial, nuanced factors driving user acceptance in various domains, especially healthcare. While SARs offer capabilities in immersion, engagement, and dynamic interaction tablets retain a significant advantage rooted in established user familiarity and inherent simplicity of operation (61,62).

The TAM provides a framework for dissecting these acceptance variances, as PU and PEU are predictably heterogeneous across these technology types. Conversely, SARs often suffer from a lower PEU score due to the steeper learning curve associated with new interaction paradigms (e.g., gesture control, spatial calibration) and the novelty of the immersive experience itself. In contrast, tablets, being ubiquitous, benefit from near-zero PEU barriers for most users (62).

Moreover, the strategic role of intrinsic and extrinsic motivational elements within the user experience cannot be overstated. SARs often integrate highly effective motivational scaffolding, such as personalized, context-aware feedback. Studies suggest that such rich, personalized feedback mechanisms significantly boost the perceived value and sustained engagement with SARs, particularly within therapeutic or educational contexts where adherence and sustained motivation are paramount for successful outcomes (26,65,68). This difference in motivational scaffolding directly impacts long-term adoption rates. Therefore, understanding these intricate dynamics is essential for the strategic optimization of the design, deployment, and ultimate clinical efficacy of both SAR and tablet-based solutions in diverse healthcare settings.

1.4 Research gap

The adoption of SARs in healthcare represents a rapidly expanding field of research and practice, driven by the need for innovative tools to enhance patient care and address workforce shortages (26). While SARs have demonstrated significant potential in various applications, including cognitive training, physical rehabilitation, and emotional support, there are only a few studies conducted in clinical psychiatric settings. The use of SARs in adult psychiatric settings has been investigated far less extensively than in other areas of care; however, emerging

evidence suggests promising approaches. In a scoping review, Kling et al. (69) identified studies addressing various psychiatric and neurodevelopmental conditions, including schizophrenia, autism spectrum disorder, and intellectual disabilities. In these studies, SARs were employed to reduce specific symptom dimensions, such as negative symptoms in schizophrenia, to promote social skills, for example through simulated job interviews for individuals with ASD, and to enhance engagement and participation among people with intellectual disabilities.

Complementing these findings, an umbrella review by Nichol et al. (70) reported consistent evidence across the lifespan for beneficial effects of SARs on loneliness, social interaction, mood, and positive effect. Notably, in group-based settings, robots appeared to facilitate interpersonal interaction rather than replace it, underscoring their potential role as supportive tools within psychosocial interventions.

In contrast, the evidence regarding the reduction of specific psychiatric symptoms, such as depression, anxiety, or agitation, remains heterogeneous. While meta-analyses in older adult populations did not demonstrate a significant overall effect on depressive symptoms or agitation, individual primary studies reported positive trends in these domains (70). In children, however, a consistent reduction in anxiety symptoms associated with the use of socially assistive robots was observed, suggesting potential age- and development-related differences in intervention effectiveness (70).

SARs have been studied in geriatric care, with evidence indicating their effectiveness in fostering patient engagement, improving treatment adherence, and enhancing outcomes such as reducing loneliness and promoting social interaction. These findings highlight their potential to provide meaningful emotional and psychological support (71–76).

In rehabilitation settings, SARs have been employed to motivate patients during physical therapy, often outperforming traditional methods by sustaining engagement and improving task performance (77).

Studies in caregiving contexts demonstrate that SARs often outperform traditional interventions, such as tablet-based training, in engaging users and improving outcomes (78). A systematic review by Chen et al. (27) highlighted the potential of SARs in reducing depression among older adults. In several studies included in the review, interventions with SARs like "Paro" significantly reduced depressive symptoms, loneliness, and improved quality of life. Qualitative interviews revealed that participants anthropomorphized the robots, forming personal connections that enhanced their psychological well-being (79).

Another study employing the social animal robot "PIO" found significant improvements in cognitive function, depressive symptoms, and loneliness among older adults participating in a cognition-based intervention program (80). The findings underscore the unique ability of SARs to combine cognitive training with emotional and social support, addressing the multifaceted challenges of depression. A systematic review by Duradoni et al. (81) identified three psychological domains – social skills, mood, and well-being – that benefit from SAR-based interventions. For instance, the humanoid robot "Ryan" demonstrated potential in delivering cognitive-behavioral therapy, improving depressive symptoms in individuals with mild to moderate depression (82). These studies highlight the versatility of SARs in addressing both cognitive and emotional dimensions of mental health.

In pediatric care, SARs have been used to support children with autism spectrum disorders (ASD), where they have shown promise in improving social communication skills and reducing anxiety during medical procedures (83). These applications highlight the versatility of SARs in addressing diverse healthcare needs across populations and settings. However, the generalizability of these findings to psychiatric care, particularly for adult populations, remains unclear due to fundamental differences in the nature of psychiatric disorders and the unique challenges faced by this demographic (84–86).

Research on SARs in psychiatric settings is relatively limited compared to other healthcare domains. Existing studies have primarily focused on pediatric populations, such as children with ASD or attention deficit hyperactivity disorder (ADHD), where SARs have been shown to provide valuable therapeutic benefits (83,87–90). For example, robots like NAO and Kaspar have been used to facilitate structured interventions aimed at improving emotional regulation and social skills in children, often with positive outcomes (83). However, the use of SARs in adult psychiatric care has received far less attention, despite the growing prevalence of mental health disorders among adults worldwide (91).

The lack of research on socially assistive robots in adult psychiatric settings is particularly concerning given the distinct needs and characteristics of this population. Adults with psychiatric disorders, such as depression, schizophrenia, and anxiety disorders, often face complex challenges, including cognitive impairments, social withdrawal, and emotional dysregulation, which can substantially interfere with daily functioning, treatment adherence, and sustained engagement with digital or technology-based interventions (92,93). These impairments may limit patients' ability to interact consistently with technology-driven interventions, follow structured therapeutic programs, or maintain motivation over time, underscoring the importance of tailored and supportive technological solutions in psychiatric

care. SARs have the potential to address these issues by providing personalized cognitive training, fostering social interaction, and delivering motivational support. However, the effectiveness of SARs in achieving these goals within adult psychiatric populations remains largely unexplored.

One of the key gaps in literature is the lack of differentiation between the effects of SARs on children and adults. While studies have shown that SARs can positively impact children with psychiatric disorders, it is not appropriate to generalize these findings to adults. Children and adults differ fundamentally in their cognitive, emotional, and social development, which influences how they perceive and interact with SARs (94). For instance, children may view SARs as playful companions and respond positively to their anthropomorphic features, whereas adults may require more sophisticated interactions and functionalities to perceive SARs as credible and useful tools (26,95).

Furthermore, the developmental stages of children and adults influence their respective therapeutic needs and outcomes. Children are more likely to benefit from interventions targeting developmental milestones, such as social communication and emotional regulation, whereas adults may require interventions that address established cognitive deficits and functional impairments (28,96). The lack of studies examining these differences limits the ability to tailor SAR interventions to the specific needs of adult psychiatric populations.

Another critical research gap pertains to gender and sex differences in the acceptance and effectiveness of SARs in psychiatric care. While some studies have explored gender and sex differences in technology acceptance more broadly, little is known about how these differences manifest in the context of SARs (53). Gender may influence perceptions of SARs, with women potentially valuing their relational and interactive features more highly, while men may prioritize their functional and task-oriented aspects (97). Understanding these dynamics is essential for designing SAR interventions that are tailored to the preferences and needs of different user groups.

Gender differences may also affect the outcomes of SAR-based interventions. For example, previous research has suggested that women are more likely to engage with and benefit from motivational feedback, which is a key feature of SARs (98). Investigating these gender- and sex-specific effects in adult psychiatric populations could provide valuable insights into how SARs can be optimized to enhance their effectiveness and accessibility.

1.5 Research aims, research questions and hypotheses

This chapter describes the research aim on the one hand and the specific research question on the other.

The primary aim of this pilot study is to evaluate the effects of cognitive training sessions conducted with the SAR “Pepper” on

1) depressive symptoms,
2) attention and performance,
3) acceptance and usability, and
4) motivation,

compared to cognitive training sessions conducted with a tablet, in participants with depression during acute psychiatric inpatient treatment, and whether the effects differ by sex and age.
An additional aim is to explore the participants’ experience of cognitive training with the SAR "Pepper".

The following research questions can be derived from this:

- What are the effects of cognitive training sessions delivered via the socially assistive robot “Pepper” compared to those delivered via a tablet on depressive symptoms, attention and performance, acceptance and usability, and motivation in adults with depression during acute psychiatric inpatient treatment and how do sex and age differences influence these effects?
- How do participants in the intervention group experience cognitive training with the SAR "Pepper"?

Based on the study objectives, the following null hypotheses were formulated:

H0-1 – Depressive symptoms:
There is no significant difference in change in depressive symptoms from pre- to post-intervention between participants receiving cognitive training with the SAR “Pepper” and those receiving tablet-based cognitive training; furthermore, effect on depressive symptoms does not differ by sex or age.

H0-2 – Attention and performance:

There is no significant difference in change in attention and performance from pre- to post-intervention between participants receiving cognitive training with the SAR “Pepper” and those receiving tablet-based cognitive training; furthermore, effect on attention and performance does not differ by sex or age.

H0-3 – Acceptance and usability:

There is no significant difference in acceptance and usability outcomes between participants receiving cognitive training with the SAR “Pepper” and those receiving tablet-based cognitive training; furthermore, effect on acceptance and usability does not differ by sex or age.

H0-4 – Motivation:

There is no significant difference in motivation outcomes between participants receiving cognitive training with the SAR “Pepper” and those receiving tablet-based cognitive training; furthermore, effect on motivation does not differ by sex or age.

2. Methods and Material

This chapter outlines the methods and materials used in this study to provide a comprehensive overview of the methodological framework and resources used in this research.

2.1 Study design, setting and sample

This pilot study employed a randomized controlled parallel two-arm design (RCT) conducted in accordance with the CONSORT 2025 Statement (99) to ensure methodological rigor and transparent reporting. The trial was implemented between June and October 2024 at the Clinical Division of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Austria. The study sample comprised 32 inpatients diagnosed with depression by their treating clinicians based on ICD-10 criteria. The severity of depressive symptomatology was additionally assessed using the Montgomery-Åsberg Depression Rating Scale (MADRS) by a trained investigator (100,101) and by the patients themselves, using the the Beck Depression Inventory–Revised (BDI-II) (102).

2.2 Recruitment and informed consent of the participants

The recruitment of inpatients was conducted by a member of the study team, who provided a detailed explanation of the study's objectives, procedures, and potential benefits and risks. Participants were given enough time to ask questions and consider their participation before providing written informed consent. Once the informed consent form was signed, each participant was assigned a unique identification code to ensure anonymity and facilitate the pseudonymization of data.

Following the assignment of the participant code, an appointment for the first cognitive training session was scheduled. This process ensured a smooth transition into the study while allowing participants to prepare for their initial session. The systematic recruitment and onboarding procedures aimed to foster trust, ensure compliance with ethical standards, and maintain consistency in participant engagement.

2.3 Inclusion & exclusion criteria

Table 1 provides a list of all the inclusion and exclusion criteria for the RCT.

Table 1: *Inclusion and exclusion criteria*

Inclusion criteria	Exclusion criteria
Adult women and men (18-60 years)	Patient refuses to participate
Medical diagnosis of depression (ICD-10)	Diagnosed severe personality disorder
Speak and understand German	Psychotic symptoms
Can give consent	Brain organic diseases and dementia
Have no physical limitations that preclude the use of a robot	Clearly substance-induced clinical picture
	Reduced intelligence (IQ < 70)
	Secured hospital unit

ICD-10 = International Statistical Classification of Diseases and Related Health Problems Version 10; **IQ** = Intelligence quotient

2.4 Sample size calculation

The sample size estimation was based on a repeated-measures AN(C)OVA design with two measurement points (η^2 = 0.3, power = .90, α = .05), resulting in a target sample of N = 32 participants (103). This design was selected because the broader “AMIGA” (FFG grant No. FO999892829; (104)) project comprised several pre–post assessments. As a pilot study, the “AMIGA” project aimed to provide preliminary insights into depression symptoms, attention and performance, usability and acceptance and motivation of SAR-supported cognitive training in individuals with depression. The sample size calculation was performed using G*Power software (version 3.1.9.6) (105,106).

2.5 Randomization and Blinding

Upon obtaining informed consent from the participants, group allocation was conducted using the online randomization tool “StudyRandomizer.com” (107). This approach ensured an unbiased assignment of participants to either to SAR “Pepper” group with motivational feedback (= intervention group) or the tablet group without motivational feedback (= control group). The randomization process was carefully stratified to achieve balance between the groups with respect to (a) sex, (b) severity of depression measured with the MADRS score at admission, and (c) age (≤ 30 years vs. ≥ 31 years).

Age was incorporated as a stratification variable during randomization to control for potential confounding effects. The decision to categorize age into two groups, rather than using multiple

strata, was primarily driven by sample size and statistical power considerations. Given the limited number of participants, using more fine-grained age categories would have resulted in very small subgroup sizes, thus reducing the reliability and interpretability of subgroup comparisons. Dichotomization, therefore, represented a pragmatic approach to maintain analytical feasibility while still allowing for the exploration of age-related effects (108). The threshold of 30 years was selected based on theoretical and empirical considerations related to cognitive development and technology familiarity. Research indicates that patterns of digital technology use, familiarity, and attitudes often differ between younger adults and individuals in later adulthood, with younger cohorts typically demonstrating higher digital affinity and adaptability to novel technologies. The 30-year cutoff is employed in psychosocial and technology-related research as a pragmatic demarcation between early adulthood and later adult life stages (109–111).

These factors were selected to minimize potential confounding variables and ensure the comparability of baseline characteristics across the groups.
Despite the rigorous randomization process, blinding was not feasible in this study due to the inherent visibility of the intervention methods. Participants and researchers could easily distinguish between the socially assistive robot and the tablet-based intervention.

2.6 Study course

This section describes the exact procedure that the patients went through. To ensure that the study procedure was be carried out in quiet and peace, it was carried out in a specially designated room (“RoboLab”) under the supervision of a trained project member.
Participants were randomly assigned to one of two groups: (1) the intervention or (2) the control group. The study spanned a period of five consecutive days, during which participants took part in two cognitive training sessions using the "MMA-App" (Multimodal Activation-App). This structured approach aimed to evaluate the effectiveness of the intervention in comparison to the control condition.
The study commenced with participants completing baseline questionnaires (t_0) through the LimeSurvey platform (107), which provided an efficient and standardized method for data collection. Following the completion of the baseline assessment, participants engaged in their first cognitive training session. This session not only familiarized participants with the MMA-App and robot “Pepper” but also established the foundation for subsequent cognitive exercises.

At the end of the first session, participants scheduled their second session, ensuring adherence to the study timeline and minimizing potential attrition.

The second session mirrored the structure of the first. Participants underwent another round of cognitive training, reinforcing the skills and exercises introduced earlier. Upon completing the training, participants were prompted to fill out the post-intervention questionnaires (t_1), again facilitated through LimeSurvey (107). This final data collection marked the conclusion of their participation in the study.

Each complete session required, on average, between 45 and 60 minutes, offering a balance between cognitive engagement and administrative tasks. Of this time, approximately 10 to 20 minutes were dedicated to the cognitive training component, while the remaining time was used for completing questionnaires and arranging subsequent sessions. The course of the study and the adapted patient flow diagram (99) is shown in **Figure 3**.

2.6.1 Intervention group

In the intervention group, participants engaged with integrated “MMA-App” on the tablet of SAR “Pepper”. This combination allowed for an interactive and supportive cognitive training experience. The training consisted of three structured tasks designed to challenge and improve cognitive functions. Upon completion of each task, SAR “Pepper” provided motivational feedback to reinforce the participants' efforts and encourage continued engagement with the training.

The robot’s feedback was tailored to create a positive and supportive environment, fostering a sense of accomplishment and motivating participants to persevere through training. This interactive component was intended to enhance the effectiveness of cognitive training by addressing not only cognitive but also emotional and motivational aspects of the intervention. By integrating technological functionality with human-like interaction, SAR “Pepper” aimed to create a novel and engaging therapeutic experience for participants (see **Figure 2**).

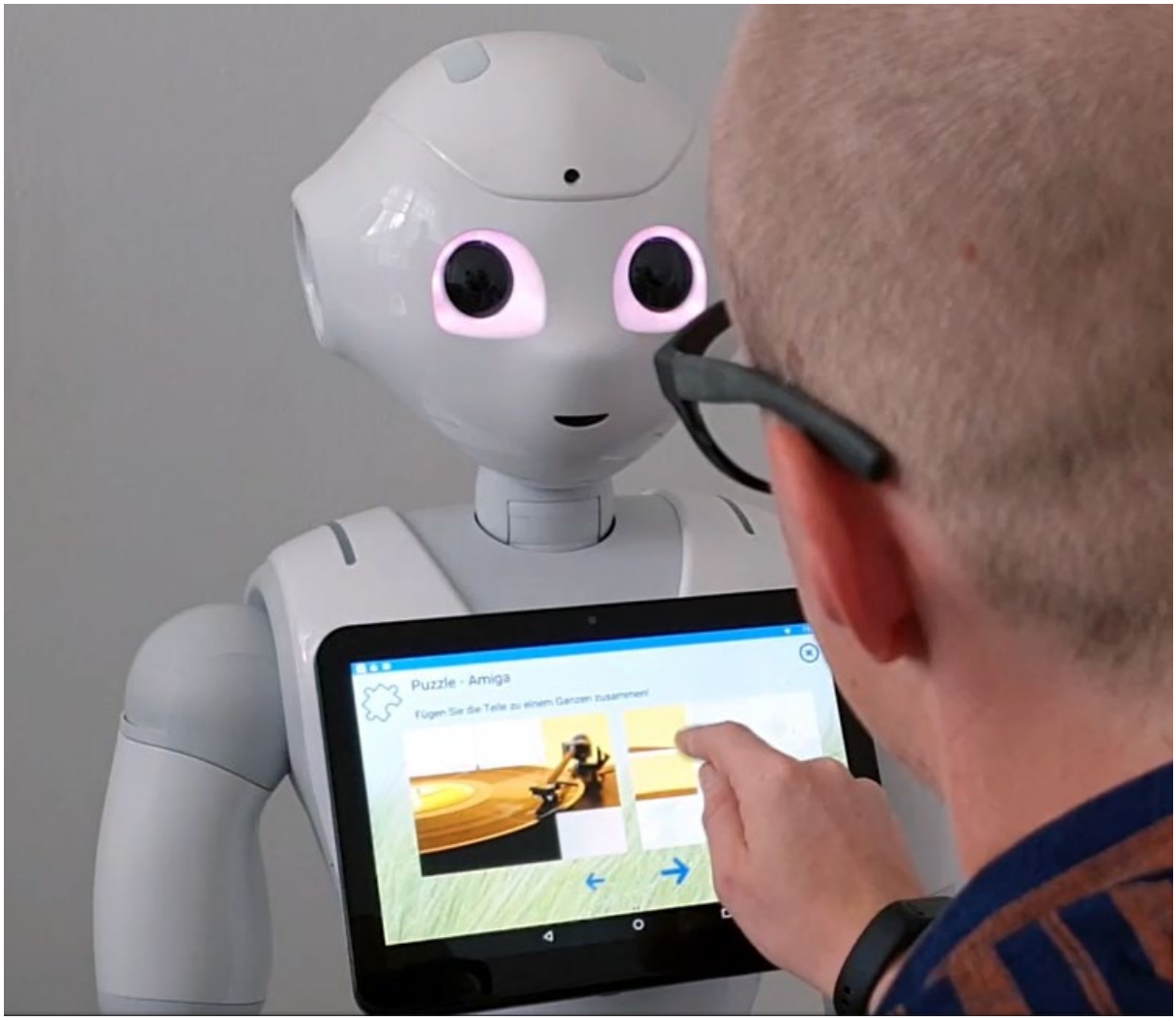

Figure 2: *Participant interacts with the tablet-supported MMA application on the SAR "Pepper" and completes the cognitive task (jigsaw) (The participant provided written informed consent for this photograph to be taken and for its publication; Credits: Sandra Draxler – JOANNEUM RESEARCH DIGITAL)*

2.6.2 Control group

Participants in the control group underwent the same cognitive training exercises as those in the intervention group; however, the training was delivered exclusively through the "MMA-App" on an Android tablet. The use of the app alone ensured that the cognitive tasks remained identical across both groups, allowing for a direct comparison of outcomes based solely on the mode of delivery.

Unlike the intervention group, the control group did not receive motivational feedback during the training sessions. The absence of interactive support, such as that provided by SAR "Pepper", served as a key distinction between the groups. This differentiation allowed the study to isolate and assess the potential added value of SAR Pepper's motivational interaction in enhancing engagement and training outcomes.

Aside from the delivery method and the presence of motivational feedback, all other procedures were consistent between the two groups, including the cognitive tasks, session structure, and data collection protocols. This uniformity ensured that any observed differences in outcomes could be attributed to the mode of delivery rather than other procedural variations.

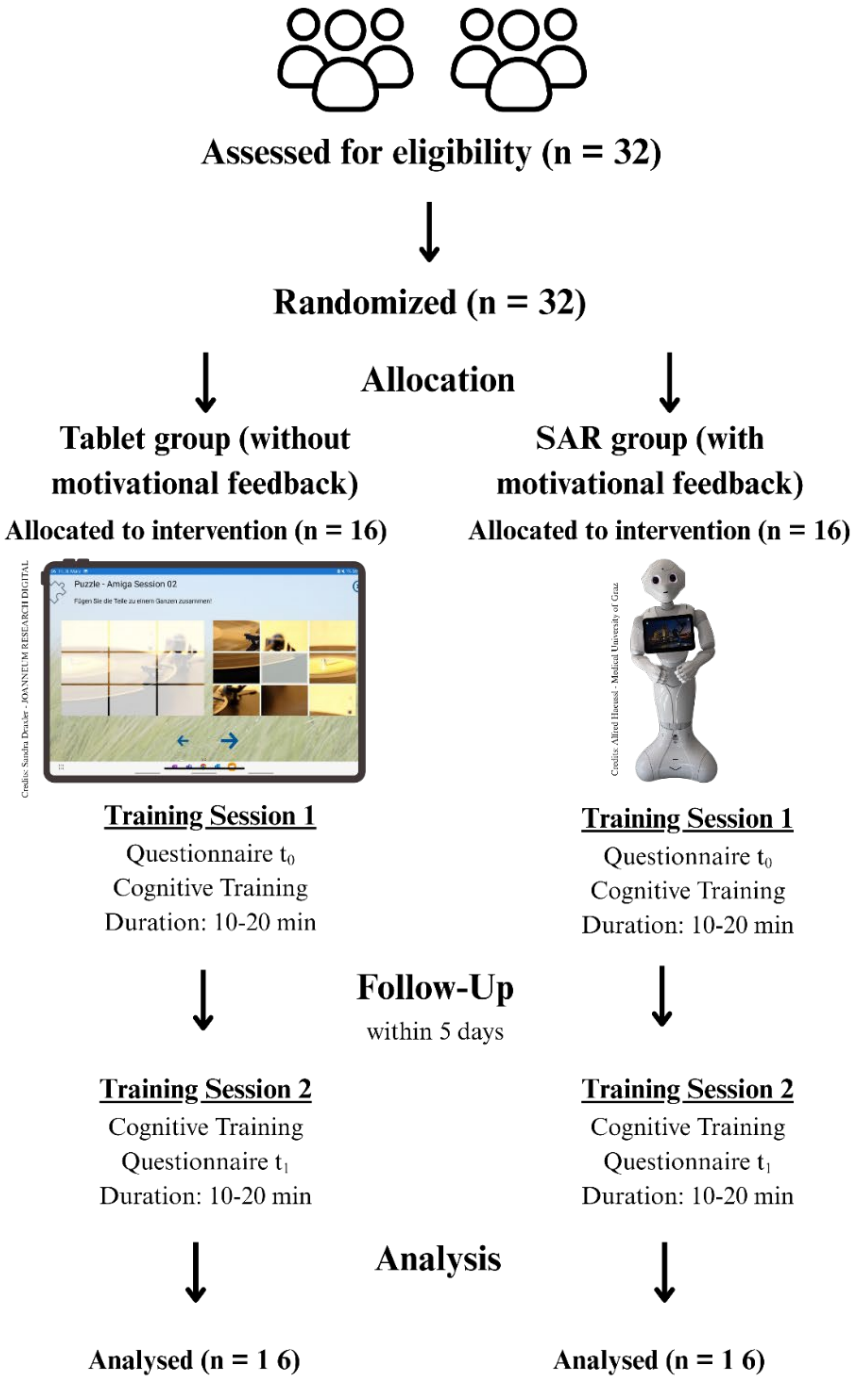

Figure 3: *Study course and adapted patient flow diagram by Hopewell et al. (99) created by Alfred Haeussl. Reproduced from Haeussl et al. (112) with permission of publisher (Frontiers in Psychiatry – Digital Mental Health).*

2.7 Pseudonymization of the data

To ensure data confidentiality, a unique participant code was assigned to each participant, serving as the sole identifier for all study-related data. These codes were used to encrypt the data, safeguarding participant anonymity throughout the study. Access to the list linking participant codes to personal identifiers was strictly limited to study team members and stored in a password-protected document to prevent unauthorized access.

All data files were labeled exclusively with participant codes, further reinforcing privacy. The encrypted database was securely stored on a password-protected server, with access restricted to authorized study team members only. These measures adhered to strict data protection standards, ensuring compliance with ethical guidelines and regulatory requirements for handling sensitive information in clinical research.

2.8 Material

2.8.1 SAR "Pepper"

Photo Credits: Alfred Haeussl, Medical University of Graz

SAR "Pepper", a humanoid robot developed by SoftBank Robotics, stands 120 cm tall and weighs 28 kg (see **Figure 4**). Its design enables sophisticated interactions through movements of its arms, hands, and head, creating a dynamic and engaging presence. SAR "Pepper" is equipped with advanced capabilities to analyze and respond to human emotions, using cues such as body posture, speech patterns, facial expressions, and vocal tone. These features enable SAR "Pepper" to provide a highly interactive and responsive experience. With integrated microphones, SAR "Pepper" can identify and orient itself toward the speaker, fostering a natural and intuitive form of communication (113–115).

SAR "Pepper" is specifically designed to enhance human-robot interaction in therapeutic, rehabilitative, and educational contexts. Its engaging and friendly demeanor is aimed at increasing motivation and improving the overall quality of life for users (113–115). Research underscores the potential of SARs like "Pepper" to provide meaningful support in various domains, from promoting adherence in therapy to facilitating learning processes (28). SAR

“Pepper’s” unique combination of advanced technological features and human-centered design makes it a valuable tool in environments requiring social interaction and support (28).

2.8.2 “MMA-App” and Cognitive Training

The “MMA” (multimodal activation) application developed Joanneum Research – Digital (116,117), comprises a tablet-based front end and a back end hosted on a central server. Both components were built using the Microsoft “.NET Framework”. The front end was developed with Xamarin, a cross-platform framework by Microsoft that supports mobile application development in “.NET”. The application was specifically designed for Android devices and optimized for tablet displays. To support user management and data exchange, an application programming interface (API) was implemented, enabling authentication, version updates, and the transmission of exercise results. A structured query language (SQL) database was created to store user information and collected results. Additionally, a content management system (CMS) was integrated to facilitate the creation and thematic categorization of various exercise types into cohesive content units (75,118).

The “MMA-App” features a serious game that has been specifically developed and tested for people with dementia (116,117). This multimodal training tool combines cognitive and physical exercises in an engaging and playful format. The activities are highly customizable to suit individual needs, including adjustments to content, difficulty level (based on cognitive abilities), sequence, and duration. The training sessions in this study only consist of cognitive tasks, such as quizzes, puzzles, memory matching games, or arithmetic problems. “MMA-App” is designed for flexible use, allowing it to be applied both at home and in institutional settings (see **Figure 5**).

The cognitive training sessions were structured into three distinct exercise blocks, each containing six tasks, resulting in a total of 18 tasks per session. To maintain engagement and progressively challenge participants, the difficulty level increased after the completion of each block. The first block began at difficulty level 2 out of 4, the second block advanced to level 3 out of 4, and the final block reached the highest difficulty level, 4 out of 4. This adaptive design aimed to promote gradual skill development and ensure participants remained motivated throughout the training.

Each session featured a selection of three cognitive tasks, carefully chosen from different cognitive domains to provide comprehensive training experience. During the first session, participants completed tasks focusing on memory and problem-solving skills: finding picture pairs (memory domain), solving arithmetic problems (working memory domain), and filling in

missing words in a text (semantic memory domain). These tasks were designed to stimulate various cognitive processes while maintaining a manageable workload for participants.

In the second session, the cognitive tasks shifted to include additional domains, reflecting a broader range of cognitive abilities. Participants solved arithmetic problems (working memory domain), completed puzzles (visuospatial memory domain), and organized a sequence of events in the correct chronological order (seriality and long-term memory domains). This varied task design not only targeted multiple cognitive functions but also minimized repetition, keeping the training sessions engaging and dynamic (see **Table 2**).

By combining structured progression, adaptive difficulty levels, and diverse cognitive tasks, the training program was designed to provide a balanced and effective approach to cognitive skill enhancement across different domains.

***Table 2**: Overview of Training Activities by Difficulty Level and Session*

Difficulty level	Training session 1	Training session 2
Level 2	Picture Pairs	Mathematics Quiz
	Picture Pairs	Mathematics Quiz
	Mathematics Quiz	Jigsaw Puzzle
	Mathematics Quiz	Jigsaw Puzzle
	Cloze Text	Sequence
	Cloze Text	Sequence
Level 3	Picture Pairs	Mathematics Quiz
	Picture Pairs	Mathematics Quiz
	Mathematics Quiz	Jigsaw Puzzle
	Mathematics Quiz	Jigsaw Puzzle
	Cloze Text	Sequence
	Cloze Text	Sequence
Level 4	Picture Pairs	Mathematics Quiz
	Picture Pairs	Mathematics Quiz
	Mathematics Quiz	Jigsaw Puzzle
	Mathematics Quiz	Jigsaw Puzzle
	Relaxation	Relaxation
	Cloze Text	Sequence
	Cloze Text	Sequence

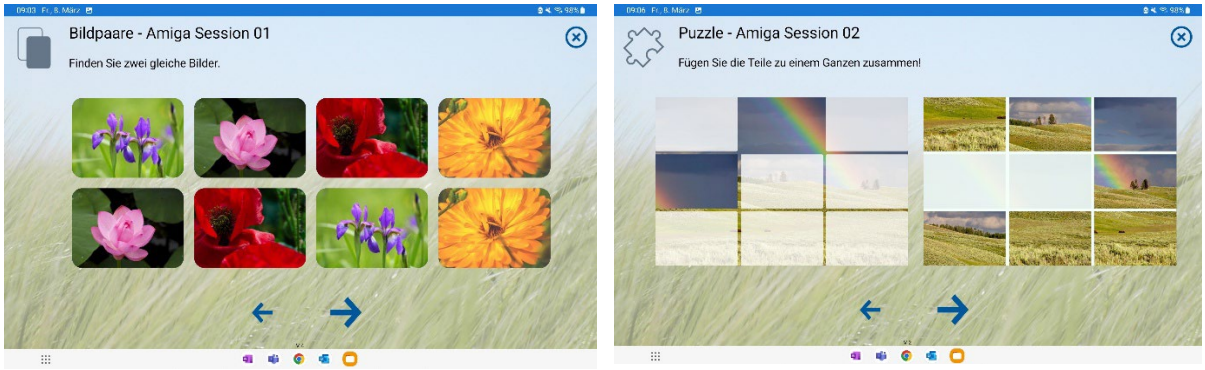

Figure 5: *Screenshots of the MMA-App (Picture Pairs & Jigsaw). Photo Credits: Lucas Paletta (Joanneum Research - Digital)*

2.8.3 Questionnaires

In the following section, all questionnaires used in this study are described in detail. Both standardized and non-standardized (self-created) questionnaires were utilized.

2.8.3.1 Standardized questionnaires

Below, all standardized questionnaires are described.

2.8.3.1.1 Montgomery–Åsberg Depression Rating Scale (MADRS)

The MADRS is a clinician-administered scale designed to assess the severity of depressive symptoms in individuals diagnosed with depression (100,101). Developed in 1979 by Stuart Montgomery and Marie Åsberg, the MADRS was specifically created to provide a sensitive tool for evaluating changes in depressive symptoms, particularly in response to treatment (100,101). It is widely used in both clinical settings and research studies due to its reliability, sensitivity, and focus on core depressive symptoms (100,101). The MADRS consists of 10 items, each representing a symptom commonly associated with depression, such as sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. Each item is rated on a 7-point scale, ranging from 0 (no symptom) to 6 (severe symptom)**.** The total score is calculated by summing the ratings for all items, yielding a possible range of 0 to 60 (100,101). Higher scores indicate greater severity of depressive symptoms, with thresholds often interpreted as follows:

- 0 to 6 = no depression
- 7 to 19 = mild depression
- 20 to 34 = moderate depression
- 35 to 60 = severe depression.

One of the key advantages of the MADRS is its focus on the emotional and cognitive aspects of depression, rather than somatic symptoms, making it particularly useful for detecting treatment-related changes. This is especially important in pharmacological trials and other therapeutic interventions where the primary aim is to evaluate symptom reduction over time (100,101). The scale's sensitivity to change is one of its most valued attributes, making it a preferred tool for monitoring treatment outcomes. MADRS has demonstrated strong psychometric properties. It exhibits high reliability, with consistent results across different raters, and good validity, as it correlates well with other depression rating scales, such as the Hamilton Depression Rating Scale (HAM-D) (100,101). Furthermore, the MADRS is relatively quick and easy to administer, typically taking 15 to 20 minutes, which contributes to its widespread use in clinical and research settings. Despite its strengths, the MADRS has limitations. As a clinician-administered scale, it requires trained personnel to ensure accurate and consistent ratings. Additionally, while the MADRS is effective in measuring the severity of depressive symptoms, it does not provide a comprehensive assessment of all facets of depression, such as functional impairment or comorbid conditions (100,101).

2.8.3.1.2 Beck Depression Inventory Revision (BDI-II)

The BDI-II is a widely used self-report questionnaire designed to assess the severity of depressive symptoms in individuals aged 13 years and older. Developed by Aaron T. Beck, a pioneer in cognitive therapy, the BDI-II is an updated version of the original Beck Depression Inventory (BDI), which was first introduced in 1961 (102). The BDI-II was revised in 1996 to align with modern diagnostic criteria for major depressive disorders, as outlined in the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) (102). This revision included updates to item content, such as the addition of items addressing worthlessness, loss of energy, and irritability, as well as a change in the time frame for symptom evaluation to the past two weeks instead of the current moment (102). These changes were made to enhance the tool's relevance and alignment with clinical standards. The BDI-II consists of 21 items, each corresponding to a specific symptom of depression, such as sadness, loss of pleasure, sleep disturbances, and appetite changes. Respondents rate the severity of each symptom on a 4-point Likert scale, ranging from 0 (absence of the symptom) to 3 (severe symptom) (102). The total score is calculated by summing the ratings across all items, with possible scores ranging from 0 to 63. The total score provides an indication of the severity of depressive symptoms, with established thresholds as follows (102).

- 0 to 8 points = no sign of depression
- 9 to 13 points = minimal depression
- 14 to 19 points = mild depression
- 20 to 28 points = moderate depression
- 29 to 63 points = severe depression

It is important to note that BDI-II is not intended to provide a clinical diagnosis but rather to measure symptom severity and track changes over time. BDI-II is widely applied in clinical practice, research, and educational settings (102).

The tool has strong psychometric properties contribute to its popularity. It demonstrates high reliability, with internal consistency (Cronbach's alpha) typically exceeding 0.90, and good validity, as it correlates well with other measures of depression and psychological distress (102).

2.8.3.1.3 Inventory of Depression and Anxiety Symptoms – German Version (IDAS-II)

The IDAS-II is a comprehensive self-report measure designed to assess a wide range of symptoms related to depression and anxiety (119). The IDAS-II comprises 98 items, which are grouped into 18 specific subscales. These subscales measure distinct aspects of depression and anxiety, including core depressive symptoms (e.g., Dysphoria, Suicidality), anxiety-related symptoms (e.g., Panic, Social Anxiety), and additional associated factors, such as appetite changes, insomnia, and irritability. Respondents rate the severity of each symptom on a 5-point Likert scale ranging from 1 (not at all) to 5 (extremely), based on their experiences over the past two weeks (119). One of the key strengths of the IDAS-II is its ability to differentiate between symptoms of depression and anxiety, which often co-occur and can be difficult to distinguish using broader measures (119).

A “Depression” Scale can also be constructed, combining all items from the “Dysphoria” scale with two selected items from the scales “Suicidality”, “Fatigue”, “Insomnia”, “Appetite Loss”, and “Well-being”. This specific combination was designed to ensure alignment with established depression questionnaires, such as the BDI-II. The General Depression Scale includes the following items: 1, 2, 5, 6, 8, 9, 11, 13, 21, 26, 27, 30, 31, 39, 47, 50, 51, 56, 60, 63 (119).

The IDAS scale is also available in German (Wester et al., 2021). The IDAS-II demonstrates strong psychometric properties, making it a reliable and valid tool for assessing symptoms of depression and anxiety. In terms of reliability**,** the IDAS-II exhibits high internal consistency across its scales, with ω values exceeding 0.80 for all scales, except for the “Euphoria” (*ω*

=.75) and "Compulsion" (ω = .73) subscales, which still reflect acceptable levels of reliability (119,120). Regarding validity**,** the IDAS-II shows robust evidence for convergent and discriminant validity**.** It correlates strongly with other established measures of depression and anxiety, demonstrating its ability to accurately capture symptoms related to these constructs (convergent validity). At the same time, the IDAS-II effectively differentiates between related but distinct symptom dimensions, highlighting its capacity to provide nuanced assessments that distinguish depression from anxiety symptoms (discriminant validity) (119,120).

Despite its strengths, the IDAS-II has limitations. The length of the questionnaire may pose a challenge in settings where time is constrained, and the self-report nature of the instrument may introduce biases related to self-awareness or social desirability. Additionally, while the IDAS-II provides a detailed assessment of symptom severity, it does not evaluate functional impairments or broader psychosocial factors that may contribute to emotional distress (119,120).

2.8.3.1.4 Attention and Performance Self-Assessment (APSA)

The APSA is a self-report questionnaire designed to evaluate and analyze everyday memory and concentration abilities in individuals with various subclinical or clinically significant underlying conditions. The APSA consists of 21 items and provides results in the form of an overall score (APS-20) and two subscale scores: "Difficulties Prospective Memory Performance" (AP-F1) and "Difficulties in Maintaining Focused Attention" (AP-F2)**.** Each subscale is derived from nine specific items, making the APSA a comprehensive tool for assessing cognitive and attentional functions in everyday contexts (121).

The APSA exhibits strong reliability. Internal consistency, as measured by Cronbach's alpha, is greater than .89 across all scales, indicating excellent internal consistency. Additionally, test-retest reliability, measured through intraclass correlations (ICC), shows high stability over time, with an ICC of .90 for the overall scale, .91 for AP-F1, and .87 for AP-F2. These metrics highlight the robustness and consistency of the APSA in repeated measurements (121). The validity of the APSA is supported by findings on both convergent and discriminant validity. The APSA overall score correlates as expected with related constructs, such as the subscales of the Hospital Anxiety and Depression Scale (HADS): Anxiety (r = .63) and Depression (r = .54) (121).

The APSA is particularly useful for identifying deficits in memory and attention performance, especially in populations with neurological or psychological conditions that may impact

everyday functioning. Its two subscales allow for a more nuanced understanding of cognitive challenges: AP-F1 focuses on prospective memory, the ability to remember to perform tasks in the future, while AP-F2 assesses sustained attention, highlighting difficulties in maintaining focus over time (121).

2.8.3.1.5 Technology Acceptance Model Questionnaire (TAM)

The TAM is a widely used framework for understanding and predicting the adoption and usage of technology. TAM is designed to evaluate how individuals perceive and interact with a given technological system, focusing on factors that influence their willingness to use it (51). The TAM questionnaire is composed of 14 items, divided into three key constructs: "Perceived Usefulness" (PU)," Perceived Ease of Use" (PEU), and "Intention to Use" (ITU). These constructs form the foundation of the model, enabling researchers to assess the likelihood of technology acceptance (51). Specifically, TAM includes six items to measure PU, which reflects the degree to which a person believes that using a particular system will enhance their performance or productivity. Similarly, six items evaluate PEU, defined as the extent to which an individual perceives the system as being easy to learn and use. The final construct, ITU, is assessed using two items, capturing an individual's likelihood of adopting the technology in the future (51).

To ensure relevance across various contexts, the TAM questionnaire allows for customization (51). For instance, in its original design, questions related to PU often referred to improvements in job performance, with phrases such as "in my job." However, this study omitted such job-related phrases, as they were not applicable to the research context. Similarly, the manual specifies replacing generic terms like "[this product]" with the name of the specific technology under evaluation. In this case, the term was adapted to reflect the SAR "Pepper" (51). This flexibility in wording helps tailor the TAM to specific systems and environments while maintaining its validity and reliability (51). The TAM framework has been extensively validated and is widely recognized for its predictive power and flexibility in assessing diverse technologies. By measuring PU, PEU, and ITU, Technology Acceptance Model provides actionable insights into factors that influence user acceptance, making it a valuable tool for understanding and improving the adoption of new technologies in various fields, including healthcare, education, and workplace innovation (51).

2.8.3.1.6 System Usability Scale (SUS)

The SUS is a widely used questionnaire designed to evaluate the usability of technology. Created by Brooke in 1995, the SUS is recognized for its simplicity, flexibility, and technology-independent design, making it applicable across various domains and systems. Its short length and ease of administration contribute to its popularity as a usability assessment tool. The SUS consists of ten items, each rated on a 5-point Likert scale ranging from 0 (don't agree at all) to 4 (agree completely) (122–125).

The SUS provides a single composite score ranging from 0 to 100, reflecting the overall usability of a system. To calculate the score, contributions for odd-numbered items (1, 3, 5, 7, 9) are determined by subtracting 1 from the selected scale position, while contributions for even-numbered items (2, 4, 6, 8, 10) are calculated as 5 minus the scale position. These values, each ranging from 0 to 4, are summed and multiplied by 2.5 to yield the final SUS score. This standardized process ensures a concise and interpretable measure of usability (122,123). The SUS score ranges from 0 to 100, with 68 or higher considered indicative of good usability (122). The SUS is available in 19 languages, including German, further enhancing its accessibility and applicability in international contexts (122,126). Despite its brevity and straightforward design, the SUS has demonstrated strong psychometric properties. The literature describes it as a valid scale with a high capacity to differentiate between usability levels when compared to other commercially available questionnaires (127). Its validity has been rated as high, further supporting its use in both research and practice (127).

2.8.3.1.7 Technology Usage Inventory (TUI)

TUI is a comprehensive questionnaire designed to assess various dimensions of individuals' interactions with and attitudes toward technology. It consists of 30 items, distributed across eight primary scales: "Curiosity" (CUR), "Anxiety" (ANX), "Interest" (INT), "User-Friendliness" (UF), "Immersion" (IMM), "Usefulness" (USE), "Skepticism" (SKE), and "Accessibility" (ACC). In addition, the questionnaire includes the "Intention to Use" (ITU) scale, which measures an individual's willingness to adopt the technology (128).

Each scale is rated on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), except for the "Intention to Use" (ITU) scale, which uses a visual analog scale. Respondents indicate their level of agreement by placing a mark on a horizontal line, providing a flexible and intuitive response format. A value of 0 to 100 can be achieved for each question. Most scales consist of four items, except for "User-Friendliness" and "Accessibility", and

“Intention to Use”, which contain only three items each. The flexibility of the TUI allows for tailoring questions to the specific technology being evaluated. For instance, the generic term "this technology" can be replaced with the name of the technology under study (e.g., "the socially assisting robot Pepper") to enhance clarity and relevance (128). The “Immersion” scale, designed to measure the user's sense of being absorbed or engaged with a virtual or augmented reality system, can be omitted if the evaluated technology does not fall into this category. This adaptability ensures that the TUI remains applicable across diverse technological domains, making it a versatile tool for technology acceptance research (128). The reliability of the TUI is rated as good, with internal consistency values ranging from α = .70 for the ”Skepticism” scale to α = .89 for the Interest scale, based on a sample of 178 participants (128). These findings underscore the robustness of the TUI in capturing users' attitudes and perceptions toward technology (128).

Technology Usage Inventory is a reliable and flexible instrument for assessing user attitudes, emotions, and intentions regarding a wide range of technologies. Its multidimensional structure and adaptability make it a valuable tool for researchers and practitioners seeking to understand factors influencing technology acceptance and usage.

2.8.3.1.8 Questionnaire on Current Motivation (QCM)

The QCM, originally referred to as the “Fragebogen zur aktuellen Motivation”, is a self-report instrument designed to assess an individual’s current motivation in learning and performance situations. The QCM evaluates motivational states in real-time contexts, making it a valuable tool for understanding dynamics in educational and task-based environments. It consists of 18 items distributed across four scales: “Fear of Failure”, “Interest”, “Probability of Success”, and “Challenge”. The development of the QCM was originally rooted in studies focused on learning with computers, providing a foundation for its application in technologically supported learning scenarios (129).

The reliability of the QCM is supported by internal consistency values (Cronbach’s alpha) ranging from .66 to .90, depending on the scale. These figures indicate acceptable to excellent reliability, ensuring the consistency of the measurements (129). Regarding validity, the QCM demonstrates strong factorial validity, with a four-factor structure explaining 58.4% of the variance (129). This structure aligns with the theoretical framework of the questionnaire. Additionally, the QCM shows evidence of predictive validity, as reported by Rheinberg et al. (129). These authors found that current motivation, as measured by the QCM, correlates with

subsequent learning behaviors and performance outcomes. For instance, the factors "Challenge" and "Interest" were identified as predictors of performance, although their influence depends on the specific learning task and the number of practice attempts (129). QCM is widely used in research and practice to assess motivation in educational settings, particularly when evaluating the effects of different learning environments, tasks, or interventions on motivation.

2.8.3.2 Non-standardized questionnaire

For socio-demographic and medical data, the following parameters were collected:

- Participant code
- Gender
- Age
- Marital status
- Number of children
- Highest education completed
- Main psychiatric diagnosis and other medical condition

Information about the participants' citizenship, the city where they were born, and whether they live in an urban or rural area was collected. To answer the questions, participants selected predefined response options. A number had to be entered to answer the question "age" and "number of children" and a number and letter combination had to be entered to answer, "individual code".

Additionally, the "Questionnaire on Achievement Motivation" comprised three items, each evaluated using a 7-point Likert scale. These items were designed to measure three distinct constructs related to the participants' experiences with the "BRAINMEE training"**:** "pride"**,** "shame"**,** and "confidence". Each item was analyzed independently, as the questionnaire does not allow for the calculation of a cumulative score.

The above-mentioned questions, apart from achievement motivation, were collected only once at time point t_0.

Achievement motivation was collected at both time points, t_0 and t_1. The exact questions are presented in the appendix (see **Table 15**, and **Table 16**).

At the end of the post-intervention questionnaire (t_1), participants were given the opportunity to provide additional feedback through open-ended text fields. They were invited to share their wishes, needs, suggestions for improvement, as well as positive and negative experiences related to the intervention (see **Table 17**).

2.9 Statistical analysis

This subsection describes the statistical analyses of both the quantitative and qualitative data.

2.9.1 Quantitative analysis

Demographic characteristics of the sample were examined using descriptive statistics, including the calculation of means (*M*) and standard deviations (*SD*), to provide a comprehensive overview of the participants' baseline attributes. Changes in depression (only the Inventory of Depression and Anxiety Symptoms – German Version), attention and performance and motivation accessed across two measurement points (t_0 = baseline and t_1 = post-intervention) were analyzed using repeated-measures t-tests, allowing for the assessment of within-subject differences over time.

Acceptance and usability scales were evaluated exclusively at the second measurement point (t_1). To examine group, sex and age differences in questionnaire outcomes, different statistical approaches were applied based on the structure of the respective instruments. For questionnaires consisting of a single overall scale (i.e., the System Usability Scale (SUS) and the overall score of the Technology Usage Inventory (TUI)), a one-way analysis of covariance (ANCOVA) was conducted. For instruments comprising multiple subscales (i.e., the Technology Acceptance Model (TAM) and the subscales of the TUI), a multivariate analysis of covariance (MANCOVA) was performed. The assumptions for ANCOVAs and MANCOVAs were checked and met.

To control for potential baseline differences between group, sex and age, ANCOVAs and MANCOVAs were adjusted for baseline depression severity, as measured by BDI-II t_0 and MADRS t_0 scores. Adjusted effect sizes (partial η^2) and 95% confidence intervals (CIs) are reported. To reduce the risk of Type I error due to multiple testing, Bonferroni corrections were applied to all ANCOVA and MANCOVA analyses.

Partial eta squared (η^2_p) was reported to quantify effect sizes, with values of .01, .06, and .14 representing small, medium, and large effects, respectively (130). Statistical significance was

set at $p < .05$ (two-tailed) for all analyses. The data was securely downloaded from LimeSurvey (131) and transferred to a locally stored SPSS database at the Medical University of Graz. All statistical analyses were conducted using IBM SPSS Statistics Version 30 (132), ensuring robust and precise computations. This methodological approach facilitated a detailed and reliable evaluation of the study's primary and secondary outcomes.

2.9.2 Qualitative Analysis

At the end of the post-questionnaire (t_1), several open-ended questions were integrated into the LimeSurvey (131) survey, giving participants the opportunity to provide written feedback voluntarily. A qualitative descriptive method was employed to directly describe and comprehensively summarize the subject of interest in the participants' native language, strictly adhering to the data (133,134). All written responses were copied into a separate Word document and analyzed using qualitative content analysis (135,136) using the software MAXQDA 24 (137). This method allows for a systematic and transparent interpretation of written responses by identifying recurring themes and categories. The analysis followed an inductive approach, in which categories were developed directly from the participants' responses without predefined assumptions. All statements were read several times to ensure familiarity with the data, then coded and grouped into thematic categories. To enhance reliability, the coding process was conducted independently by two researchers, and any discrepancies were discussed until consensus was reached. Representative quotations were selected to illustrate the main themes identified in the analysis.

2.10 Ethical consideration & ethics approval

The study posed no significant ethical concerns, as all procedures were carefully designed to prioritize participant safety and well-being. Eligibility was open to all inpatients receiving treatment for depression, ensuring inclusivity within the clinical population. While the sessions required active participation and could potentially lead to fatigue, participants retained the right to terminate the session at any point without any consequences. Additionally, a member of the study team was always present during sessions to address any technical issues and provide immediate assistance if needed.

The study received a positive ethics vote from the Ethics Committee of the Medical University of Graz (EK 35-450 ex 22/23), ensuring that all protocols met the highest ethical standards. The study strictly adhered to the principles outlined in the Declaration of Helsinki, emphasizing

the importance of voluntary participation, informed consent, and the protection of participant rights and data. Only participants who signed a written informed consent form were included, reinforcing the commitment to ethical integrity throughout the research process.

2.11 Funding

This study was partly funded by project AMIGA under grant No. FO999892829 of the Austrian FFG femtech programme co-financed by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK).

3. Results

This chapter provides a detailed examination of the results from the conducted pilot study. First, the basic characteristics of the study participants are described. Subsequently, the main results are analyzed and presented, focusing on comparisons between the intervention and control group, as well as sex and age differences.

3.1 Sample characteristics

This section provides a detailed overview of the study participants' baseline characteristics. Initially, the sociodemographic characteristics, offering insight into their educational background, occupation, and marital status, and the distribution of participants between the intervention and control group are described. This is followed by an analysis of the average age and depression scores at admission within these groups, broken down by sex and age strata.

3.1.1 Sociodemographic characteristics

The sociodemographic characteristics of the total sample size of 32 participants are presented in **Table 3**. Most participants (90.6%, n = 29) originated from Austria, with smaller contingents from Germany (6.3%, n = 2) and Romania (3.1%, n = 1).

Regarding education, the most common qualification was General/Middle School with A-levels, reported by 28.1% (n = 9) of participants. Apprenticeship without a master craftsman’s examination was the second most frequent, accounting for 21.9% (n = 7). Lower frequencies were reported for other education levels, such as University/Universities of Applied Sciences (9.4%, n = 3) and College (6.3%, n = 2).

Concerning current occupation, most participants were employed (56.3%, n = 18). Other categories, including unemployment (18.8%, n = 6) or working as a worker (9.4%, n = 3), were less common.

Marital status displayed a relatively even distribution between single individuals (28.1%, n = 9) and those in a relationship living in the same household (28.1%, n = 9). Fewer participants reported other statuses, such as divorced (18.8%, n = 6) and married, living together (15.6%, n = 5). No participants reported being widowed or working as farmers.

Table 3: *Sociodemographic Characteristics of Participants, including Country of Origin, Education Level, Occupation, and Marital Status.*

	Frequency	Percent
Country of origin		
Austria	29	90.6
Germany	2	6.3
Rumania	1	3.1
Highest completed education		
Compulsory School	3	9.4
Apprenticeship without master craftsman's examination	7	21.9
Apprenticeship with master craftsman's examination	1	3.1
Polytechnical School	2	6.3
General / Middle School without A-levels	3	9.4
General / Middle school with A-levels	9	28.1
College	2	6.3
Courses (e.g., universities, universities of applied sciences)	1	3.1
University or universities of applied sciences	3	9.4
Other	1	3.1
Current occupation		
Retiree	1	3.1
In Household Working	1	3.1
Civil Servant	1	3.1
Worker	3	9.4
Employee	18	56.3
Farmer	0	0.0
Self-employed	2	6.3
Not employed	6	18.8
Marital status		
Single	9	28.1
In a relationship, living in the same household	9	28.1
In a relationship, not living in the same household	2	6.3
Married, living together	5	15.6
Married, living apart	1	3.1
Divorced	6	18.8
Widowed	0	0.0

3.1.2 Distribution of participants

The sample exhibited balanced distribution across all variables. Sixteen participants were allocated to the SAR group with motivational feedback and sixteen to the tablet group without motivational feedback, ensuring an equal number in each study arm. Within both groups, the proportion of women and men was identical, with eight participants of each sex per condition. A similar pattern held for age: half of the total sample was thirty years or younger, and the other half was thirty-one years or older, again with an equal distribution of eight participants per age category in each intervention arm. Overall, the study design achieved complete balance across group, sex, and age strata, effectively minimizing the likelihood of confounding effects related to these baseline characteristics (see **Table 4**).

Table 4: *Distribution of Participants by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

	Female	Male	Overall
SAR[a]	8	8	16
Tablet[b]	8	8	16
≤ 30 Years	8	8	16
≥ 31 Years	8	8	16
Overall	16	16	32

[a] = with motivational feedback. [b] = without motivational feedback. **vs**. = versus.

3.1.3 Participants age by group and sex

The mean age of the 32 participants was 38.12 years (SD = 14.03). Analysis by group showed minimal age variation: the SAR group with motivational group averaged 37.69 years (SD = 14.59), and the tablet group without motivational feedback averaged 38.56 years (SD = 13.91). Sex analysis revealed a mean age of 37.88 years (SD = 14.90) for females and a slightly higher mean age of 38.88 years (SD = 13.48) for males. These findings indicate no significant differences in age distribution between groups or sex (see **Table 5**).

Table 5: *Descriptive statistics of participant age by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), and Sex (Female vs. Male).*

Age	*n*	*Mean*	*SD*
SAR[a]	16	37.69	14.59
Tablet[b]	16	38.56	13.91
Female	16	37.88	14.90
Male	16	38.88	13.48
Overall	32	38.12	14.03

[a] = with motivational feedback. [b] = without motivational feedback. **vs**. = versus

3.1.4 Depression scores at admission

The MADRS scores, exclusively measured at admission used as a stratification variable for randomization, (see **Table 6**) were compared between group (SAR group with motivational feedback vs. tablet group without motivational feedback), between sex (female vs. male), and age (≤ 30 years vs. ≥ 31 years).

No statistically significant group, sex or age differences were observed in MADRS scores at admission.

Table 6: *MADRS Scores at Admission by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

MADRS at Admission	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI*	
							Lower	Upper
SAR[a]	16	32.06	4.374	2.013	30	.104	-,626	6.376
Tablet[b]	16	29.19	5.282					
Female	16	30.25	4.313	2.895	30	.678	-4.401	2.901
Male	16	31.00	5.704					
≤ 30 years	16	31.06	5.221	.082	30	.621	-2.772	4.522
≥ 31 years	16	30.19	4.875					

Note: t-test. [a] = with motivational feedback. [b] = without motivational feedback.
CI = Confidence Interval. **df** = Degrees of freedom. **F** = F-statistics. **M** = Mean. **MADRS** = Montgomery-Åsberg Depression Rating Scale. **n** = Sample size of a specific group within the total sample size. **p** = p-value for t-test significance. **SAR** = Socially Assistive Robot. **SD** = Standard deviation. **vs**. = versus.

Table 7 presents the results of the BDI-II scores, measured at admission comparing the SAR group with motivational feedback and tablet group without motivational feedback, female and male participants as well as participants ≤ 30 years and participants ≥ 31 years.

Participants in the SAR group with motivational feedback reported significantly higher BDI-II scores (M = 28.25, SD = 9.61) compared to those in the tablet group without motivational feedback (M = 22.63, SD = 6.51), indicating greater depressive symptom severity at study entry in the SAR group, p = .036, 95% CI [0.49, 13.26].

All remaining comparisons across sex and age categories showed no statistically significant differences.

Table 7: *BDI-II Scores at Admission by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

BDI-II	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI*	
							Lower	Upper
SAR[a] at t_0	16	28.25	9.61	.210	30	**.036***	.494	13.256
Tablet[b] at t_0	16	22.63	6.51					
Female at t_0	16	26.00	10.02	.450	30	.482	-4.444	9.194
Male at t_0	16	23.63	10.74					
≤ 30 years at t_0	16	22.19	9.908	.335	30	.114	-11.842	1.351
≥ 31 years at t_0	16	27.44	8.278					

Note: t-test. [a] = with motivational feedback. [b] = without motivational feedback.
BDI-II = Beck Depression Inventory – revised version **CI** = Confidence Interval. **df** = Degrees of freedom. **F** = F-statistics. **M** = Mean. **n** = Sample size of a specific group within the total sample size. **p** = p-value for t-test significance. **SAR** = Socially Assistive Robot. **SD** = Standard deviation. **t_0** = Time of Measurement 1. **vs**. = versus.
* $p < .05$

3.2 Main results

3.2.1 H0-1 – Depressive symptoms

This chapter presents the results of the individual depression scales, analyzed by intervention and control groups, by sex as well as age differences. The findings are outlined for the depression subscale of the Inventory of Depressive Symptomatology (IDAS-II).

3.2.1.1 IDAS-II – Subscale Depression

Table 8 presents the scores for the Depression Subscale of the Inventory of Depression and Anxiety Symptoms-II (IDAS-II), measured at baseline (t_0) and post-intervention (t_1), comparing results by group (SAR group with motivational feedback vs. tablet group without motivational feedback), sex (female vs. male), and age (≤ 30 years vs. ≥ 31 years).

There were no statistically significant differences in IDAS-II depression scores across groups, sex, or age at either baseline or post-intervention (all $p > .05$).

3.2.2 H0-2 – Attention and Performance

This section examines attention and performance outcomes using the Attention and Performance Self-Assessment (APSA). The analysis compares scores across groups (SAR group with motivational feedback vs. tablet group without motivational feedback), sex (female vs. male) and age (≤ 30 years vs. ≥ 31 years) at baseline (t_0) and post-intervention (t_1). Results

are reported for overall APSA scores (APS-20) and its two subdimensions: AP-F1 (prospective memory performance) and AP-F2 (maintaining focused attention) (see **Table 9** and **Figure 10**)

At baseline (t_0), female participants reported significantly higher APS-20 scores (M = 67.56, SD = 10.27) compared to male participants (M = 50.63, SD = 12.11), $p < .001$, 95% CI [8.83, 25.05]. At post-intervention (t_1), female participants continued to score significantly higher on the APS-20 (M = 63.56, SD = 15.14) than male participants (M = 54.06, SD = 9.60), $p = .043$, 95% CI [0.34, 18.66].

Female participants showed significantly higher AP-F1 scores at baseline (t_0) (M = 29.25, SD = 5.96) compared to male participants (M = 20.94, SD = 5.72), $p < .001$, 95% CI [4.09, 12.53].

At baseline (t_0), female participants demonstrated significantly higher AP-F2 scores (M = 31.31, SD = 5.12) compared to male participants (M = 24.06, SD = 5.85), $p < .001$, 95% CI [3.28, 11.22]. At post-intervention (t_1), female participants continued to show significantly higher AP-F2 scores (M = 29.94, SD = 7.12) than male participants (M = 24.75, SD = 4.75), $p = .022$, 95% CI [0.82, 9.56].

All remaining comparisons, including all group and age-related analyses at baseline and post-intervention, showed no statistically significant differences.

3.2.3 H0-3 – Acceptance and Usability

This section explores acceptance and usability outcomes using three key instruments: the Technology Acceptance Model (TAM) Questionnaire, the System Usability Scale (SUS), and the Technology Usage Inventory (TUI). The analysis examines participant perceptions of the system's acceptance and usability across these measures, comparing results by intervention group and gender and sex to provide a comprehensive evaluation of user experience and technology acceptance. These three questionnaires were administered exclusively after the intervention (t_1). Consequently, depending on whether the respective instrument comprised a single or multiple subscales, either ANCOVAs or MANCOVAs were conducted. The differences between the SAR group with motivational feedback and the tablet group without motivational feedback, as well as sex-related differences (female vs. male), are reported in Häussl et al. (112) The present section focuses exclusively on age-related differences, comparing participants aged ≤ 30 years with those aged ≥ 31 years.

3.2.3.1 Technology Acceptance Model (TAM) Questionnaire

The TAM questionnaire subscales were analyzed to assess age-related differences (≤ 30 years vs. ≥ 31 years) in Perceived Usefulness (PU), Perceived Ease of Use (PEU), and Intention to Use (ITU). (see **Table 10**).

No statistically significant age-related differences were observed across any of the TAM subscales. All *p*-values remained above the Bonferroni-adjusted significance threshold (all $p > .05$).

3.2.3.2 System Usability Scale (SUS)

The ANCOVA of the SUS (see **Table 11**) revealed no statistically significant age-related differences.

3.2.3.3 Technology Usage Inventory (TUI)

Analyses of the Technology Usage Inventory (TUI) subscales revealed several statistically significant group, sex and age differences (see **Table 12**).

Participants aged 30 years or younger showed significantly higher overall technology usage (TUI OVE) scores ($M = 106.06$, $SD = 13.76$) than those aged 31 years and older ($M = 95.06$, $SD = 14.18$), $p = .018$, 95% CI [99.742, 114.511]. Younger participants also reported significantly higher interest (TUI INT) scores ($M = 20.56$, $SD = 4.24$) compared to older participants ($M = 14.75$, $SD = 4.62$), $p < .001$, 95% CI [3.328, 10.000]. Participants aged 30 years or younger additionally demonstrated significantly higher user-friendliness (TUI UF) scores ($M = 17.12$, $SD = 3.09$) compared to individuals aged 31 and higher ($M = 15.06$, $SD = 3.95$), $p = .026$, 95% CI [0.378, 5.530]. Finally, younger participants scored significantly higher on curiosity (TUI CUR) ($M = 19.88$, $SD = 5.08$) compared to older participants (M = 14.62, SD = 5.17), $p = .005$, 95% CI [1.966, 9.879].

All remaining TUI subscales showed no statistically significant differences across group, sex, or age comparisons.

3.2.4 H0-4 – Motivation

This section examines motivation using two instruments: the standardized Questionnaire on Current Motivation (QCM) and the non-standardized Questionnaire on Achievement Motivation

(QAM). The results are analyzed and presented with respect to groups (SAR group with motivational feedback vs. tablet group without motivational feedback), sex (female vs. male) and age (≤ 30 years vs. ≥ 31 years).

3.2.4.1 Questionnaire on Current Motivation (QCM)

Figure 11 and **Table 13** present the scores of the four subscales Fear of Failure (FoF), Challenge (CHAL), Probability of Success (PoS), and Interest (INT) of the Questionnaire on Current Motivation (QCM) at baseline (t_0) and post-intervention (t_1), comparing groups (SAR group with motivational feedback vs. tablet group without motivational feedback), sex (female vs. male) and age (≤ 30 years vs. ≥ 31 years).

At baseline (t_0), participants aged 31 years and older reported significantly higher QCM FoF (*M* = 3.88, *SD* = 1.48) compared to participants aged 30 years or younger (*M* = 2.65, *SD* = 1.17), *p* = .014, 95% CI [−2.204, −0.270].

At post-intervention (t_1), participants in the tablet group without motivational feedback demonstrated significantly higher QCM CHAL scores (*M* = 5.23, *SD* = 0.86) compared to those in the SAR group with motivational feedback (*M* = 4.42, *SD* = 1.18), *p* = .034, 95% CI [−1.561, −0.063].

Participants aged 31 years and older showed significantly higher baseline (t_0) QCM CHAL (*M* = 4.96, *SD* = 1.11) than younger participants aged 30 years or below (*M* = 4.09, *SD* = 1.21), *p* = .042, 95% CI [−1.714, −0.035].

All other QCM subscales showed no statistically significant differences with respect to group, sex, or age.

3.2.4.2 Questionnaire on Achievement Motivation (QAM)

The following section analyzes the results for the three subscales Pride (PRI), Shame (SHA), and Confidence (CON) of the Questionnaire on Achievement Motivation (QAM) comparing scores by group (robot vs. tablet) and gender and sex (female vs. male) at baseline (t_0) and post-intervention (t_1) (see **Table 14** and **Figure 12**).

At post-intervention (t_1), participants in the tablet group reported significantly higher pride scores (M = 5.88, SD = 1.02) compared to participants in the SAR group (M = 4.63, SD = 1.74), indicating stronger achievement-related pride in the tablet condition, p = .019, 95% CI [−2.284, −0.216].

Female participants reported significantly higher shame scores at post-intervention (t_1) (M = 5.19, SD = 2.13) compared to male participants (M = 3.44, SD = 2.06), suggesting greater self-evaluative negative affect among women, p = .025, 95% CI [0.233, 3.267].

All other subscales of group, sex and age comparisons yielded no statistically significant differences.

3.2.5 Results of the qualitative Data

In this chapter, the results of the qualitatively analyzed responses provided by the participants are presented.

3.2.5.1 General characteristics

At the end of the post-questionnaire (t_1), several open-ended questions were integrated into the LimeSurvey (131) survey, giving participants the opportunity to provide written feedback voluntarily. Therefore, not all participants submitted responses. In total, qualitative data from ten out of 16 participants in the SAR “Pepper” group (five women and five men) were included in the analysis. The mean age of these ten participants was 40 years.

3.2.5.2 General impressions and emotional reaction

Overall, participants perceived SAR “Pepper” as friendly, approachable, and non-threatening. The robot’s anthropomorphic appearance and size were often described as *“pleasantly human-like but not intimidating.”* Curiosity and mild excitement predominated during the first encounter, while a few participants initially reported unease or uncertainty about interacting with a robot. Over time, these feelings diminished as participants engaged with the system.

“At first I didn’t know how to react. But then it felt almost natural talking to him.” (P04)

“He looks kind of funny, but that makes him sympathetic.” (P10)

Emotional responses ranged from fascination to nervousness, with female participants tending to report more positive initial impressions. Participants generally appreciated Pepper's neutral design, noting that the robot was not perceived as distinctly male or female.

"It doesn't feel like a man or a woman, more like a neutral helper." (P07)

3.2.5.3 Voice and verbal communication

SAR "Pepper's" voice generated ambivalent reactions. While several participants found the tone warm and encouraging, others described it as too artificial or high-pitched, limiting emotional authenticity. The lack of voice variability was perceived as monotonous, reducing the sense of personal connection. Participants expressed a clear wish for selectable voice profiles differing in tone, speed, and gender.

"The voice was too robotic! It didn't sound real or comforting." (P01)
"If I could choose between different voices, I would probably engage more." (P06)

3.2.5.4 Movements and gestures

Participants frequently commented on the robot's physical expression. Although smooth hand gestures and nodding were viewed positively, sudden or repetitive arm movements were sometimes experienced as unsettling. Some participants even interpreted trembling hand motions as a sign of "*nervousness"* or *"malfunction."*

"When the arms move so quickly, I get a bit anxious." (P03)
"He waves too much! Once would be enough." (P09)

Nevertheless, subtle gestures accompanying motivational statements were considered beneficial for attention and engagement, as long as timing and context matched the spoken content.

3.2.5.5 Perceived Role and Usability

Despite the novelty of the humanoid interface, most participants interacted primarily with the tablet, mounted on SAR "Pepper's" chest, rather than the robot itself. SAR "Pepper" was perceived as an *"assistant"* or *"supportive figure"* rather than the main medium of training. Participants appreciated the intuitive tablet interface but criticized interruptions due to technical errors and long loading times, which occasionally broke concentration and flow.

“I mostly looked at the tablet; Pepper was more like a bystander.” (P02)
“It’s easy to use, but the waiting time between exercises makes it stressful.” (P08)

While SAR “Pepper’s” presence was described as motivational in principle, participants emphasized that malfunctions or communication delays immediately reduced their trust in the system.

3.2.5.6 Motivation and feedback

The motivational feedback provided by SAR “Pepper” elicited mixed reactions. While participants recognized the intention to encourage, the standardized phrases were perceived as repetitive and contextually misplaced. Many reported that the motivational comments were not synchronized with task performance, occasionally interrupting cognitive focus.

“Sometimes he cheered even when I did something wrong—it didn’t make sense.” (P05)
“It would help if he reacted to how I’m actually doing, not just at random times.” (P01)

A consistent suggestion was the inclusion of adaptive feedback according to task accuracy, speed, or persistence. Such individualized encouragement was seen as crucial for sustained engagement and perceived empathy.

3.2.5.7 Cognitive Training

Participants generally appreciated the cognitive training exercises as stimulating and diverse. Tasks that demanded attention, memory, and problem-solving were perceived as meaningful and challenging, though occasionally too difficult for the current phase of recovery. The absence of immediate error explanations was noted as frustrating. Several participants proposed adding a brief performance summary or the option to repeat tasks after feedback.

“I like the brain games, but I’d like to know what I did wrong.” (P04)
“Sometimes I wanted to try again, but it just moved on.” (P07)

Overall, cognitive training was rated as beneficial for structure and activation during inpatient treatment, but participants called for greater adaptability to individual cognitive capacity and daily condition.

3.2.5.8 Trust, Acceptance and Future Use

Most participants expressed openness to using SAR “Pepper” in future therapy contexts, provided that technical reliability and personalization improve. They envisioned potential applications in home-based training, outpatient care, or as a supportive element during hospitalization. Concerns mainly revolved around data protection, autonomy, and the fear of technology replacing human therapists.

“I can imagine using it at home, but only if it works without internet and saves my data locally.” (P06)

“It’s a nice addition, but it should never replace a real therapist.” (P02)

4. Discussion

The aim of this randomized controlled pilot study was to investigate the effects of cognitive training delivered via a SAR “Pepper” with motivational feedback compared to a tablet-based format without motivational feedback in adults undergoing acute psychiatric inpatient treatment for depression. Specifically, this pilot study examined outcomes related to depressive symptom severity, attention and performance, technology acceptance and usability, motivation, and subjective user experiences, while also considering sex and age differences. By combining standardized psychometric measures with qualitative feedback, this pilot study adopted a mixed-method perspective to comprehensively evaluate both clinical and experiential dimensions of technology-supported cognitive interventions in mental health care.

Across all outcome domains, we found the most consistent effects emerged for sex- and age-related differences rather than for group-based effects. Female participants demonstrated significantly higher self-reported attention and performance outcomes, including overall APSA scores as well as prospective memory and focused attention, both at baseline and partially at post-intervention.

Age-related effects were particularly evident in technology-related variables. Younger participants reported significantly higher overall technology usage, interest, user-friendliness, and curiosity, indicating greater digital affinity compared to older participants. Motivational outcomes showed selective age and group effects: older participants exhibited higher fear of failure and challenge at baseline, while at post-intervention the tablet group reported higher perceived challenge and achievement-related pride than the SAR group. Additionally, women reported higher post-intervention shame scores than men.

The present pilot study examined depressive symptom severity at admission and over the short intervention period using three established instruments (MADRS, BDI-II, and IDAS-II). The results across these measures provide important insights into baseline group differences, the comparability of demographic subgroups, and the short-term development of depressive symptoms within an acute psychiatric inpatient context.

The absence of significant group, sex, or age differences in the clinician-rated MADRS at admission indicates that all subgroups were largely comparable with respect to objectively assessed depressive severity upon study entry. This is essential for the interpretability of subsequent analyses because initial comparability minimizes the likelihood that observed post-

intervention patterns could be attributed to systematic baseline imbalances (100,101). The overall MADRS means in all subgroups were within the range typically observed among inpatients with moderate to severe depressive episodes (100,101), consistent with previous clinical samples documented in the literature (27,80,138).

In contrast to the MADRS findings, the BDI-II scores revealed a statistically significant baseline (t_0) difference between groups, with participants in the SAR group with motivational feedback reporting higher self-rated depressive symptoms than those in the tablet group without motivational feedback. This discrepancy between clinician-rated and self-rated measures has been documented in prior research and may reflect distinct perspectives captured by each instrument. Self-report instruments such as the BDI-II tend to be more sensitive to affective and cognitive components of depression, whereas clinician-administered scales often prioritize observable symptoms (139,140). Individuals experiencing depressive episodes often differ in how they subjectively perceive symptom burden, even when clinical assessments suggest comparable severity. Accordingly, the higher BDI-II scores observed in the SAR group likely reflect greater perceived psychological distress rather than fundamentally more severe clinical depression (141,142). This pattern underscores the importance of using multiple assessment modalities when characterizing depressive symptomatology in clinical trials (143).

Despite this initial imbalance, no statistically significant group differences were observed at post-intervention (t_1) for the BDI-II. Although scores remained numerically higher in the SAR group with motivational feedback, the reduction did not differ significantly between conditions. The self-rated reduction of depressive symptoms is consistent with a growing body of evidence highlighting the efficacy of structured cognitive interventions for individuals with depression (7). Cognitive training programs, such as those used in this pilot study, target core cognitive deficits frequently associated with depression, including impaired attention, memory, and executive functioning (144). By engaging participants in tasks designed to improve cognitive flexibility and problem-solving, these interventions have the potential to disrupt maladaptive ruminative thought patterns (145). The observed improvements in depressive symptoms in this study, even over a relatively short period, underscore the promise of cognitive interventions as a complementary approach to standard pharmacological and psychotherapeutic treatments (7). However, the absence of significant differences between the SAR group with motivational feedback and tablet group without motivational feedback warrants careful consideration. A key factor may be the limited duration of the intervention. The intervention in this study consisted of only two sessions over five days, which might not have been sufficient to elicit substantial differences in depressive symptomatology between groups. Previous research has shown that

longer intervention durations are often required to achieve significant and sustained reductions in depressive symptoms, as these effects typically accrue over time (146). For instance, cognitive behavioral therapy programs often span several weeks or months to allow for the consolidation of new cognitive and behavioral patterns (147).

Another explanation for the lack of significant differences could be related to the acute inpatient setting. Inpatients often experience more severe depressive symptoms and comorbidities, which may attenuate the immediate effects of interventions due to the heightened baseline symptom burden (148). Additionally, the highly structured and supportive nature of inpatient care itself may have acted as a confounding variable, providing an overarching therapeutic benefit that minimized the observable differences between the intervention groups (149).

The role of the intervention's mode of delivery also warrants attention. The inclusion of motivational feedback in the SAR group may have contributed to a more engaging and interactive experience, potentially enhancing adherence and enjoyment of the training sessions (150,151). Yet, as the content and structure of the cognitive exercises were identical across groups, it is possible that the lack of novelty in the tasks themselves limited the differentiation in outcomes (90,152). Furthermore, participants' greater familiarity with tablets compared to SARs may have influenced their initial comfort and engagement levels, favoring the tablet group in terms of perceived usefulness (51). This interplay between familiarity and novelty highlights a critical area for further exploration in future studies (153,154).

A particularly noteworthy finding in this study was the significantly higher baseline BDI-II scores in the SAR group with motivational feedback compared to the tablet group without motivational feedback. This difference, while accounted for in the statistical analysis, raises important considerations regarding the interpretation of intervention effects. Higher baseline severity of depressive symptoms can influence the capacity for improvement within an intervention group, often leading to greater variability in outcomes. Research has consistently shown that individuals with more severe depressive symptoms at baseline may respond differently to treatment compared to those with milder symptomatology, potentially due to differences in cognitive, emotional, and physiological states that affect treatment receptivity (148,150,155).

This phenomenon may be explained, in part, by the principle of "differential susceptibility," which posits that individuals with heightened emotional or cognitive vulnerability are more responsive, either positively or negatively, to interventions, depending on the quality and intensity of the treatment provided (156). In this pilot study, the SAR group with motivational feedback, characterized by higher baseline BDI-II scores, may have had a greater potential for improvement but also a higher risk of variability.

When examining sex differences in depressive symptoms, the results of this study revealed that females reported slightly higher levels of depressive symptoms at both baseline and post-intervention compared to males. However, these differences were not statistically significant, suggesting that sex may not have played a critical role in influencing depressive outcomes within the context of this study. This finding contrasts with the broader epidemiological literature, which consistently indicates that women experience higher rates of depression and more severe symptomatology compared to men (157).

No age-related differences of the BDI-II were observed at either baseline (t_0) or post-intervention (t_1). The lack of age-related differences also suggests that depressive burden at admission was relatively homogeneous across younger and older adults in this sample, which contributes to the internal validity of subsequent analyses assessing intervention effects.

The analysis of the IDAS-II depression subscale, measured both at baseline (t_0) and post-intervention (t_1), revealed no statistically significant group, sex, or age differences at either time point. This suggests that neither demographic factors nor intervention condition differentially influenced short-term changes in depressive symptoms during the study period. The absence of group differences at t_1 indicates that the brief cognitive training sessions, regardless of whether delivered via a socially assistive robot with motivational feedback or via a tablet without such feedback, did not exert immediate differential effects on depressive symptom severity. This finding is consistent with previous research indicating that cognitive training interventions typically require sustained engagement over several weeks to produce measurable improvements in mood-related outcomes (23,158). The limited number of sessions and the acute inpatient context may therefore not have provided a sufficient window for depressive symptoms to respond to the intervention.

Moreover, given the significant baseline difference in BDI-II scores between groups, it is possible that floor or ceiling effects may have constrained the extent of observable change, particularly in the SAR group with motivational feedback. Participants who reported higher subjective symptom severity at admission may have required a longer intervention period before measurable improvements could emerge. The lack of significant improvements in IDAS-II scores across groups further highlights the complexity of mood changes during acute psychiatric hospitalization. Clinical stabilization during inpatient treatment often involves multifaceted interventions, including pharmacotherapy, psychotherapy, and structured daily routines, which may overshadow small or incremental effects attributable to brief cognitive training sessions (159,160).

The congruence of null findings across sex and age categories suggests that the effects of the intervention delivery mode were generally uniform across demographic subgroups. This aligns with earlier work indicating that demographic moderators such as sex and age tend to exert limited influence on the short-term efficacy of cognitive interventions in depression, particularly when sample sizes are modest and when the intervention duration is low (161).

Given that cognitive deficits, particularly in attention and executive functioning represent a well-established feature of depressive disorders, examining performance-related outcomes provides a critical complement to the symptom-based findings presented above. The following section therefore addresses attention and performance to determine whether the cognitive training sessions produced domain-specific benefits.

In both the SAR group with motivational feedback and tablet group without motivational feedback, changes in overall APSA scores and subscale scores (AP-F1 for prospective memory and AP-F2 for sustained attention) were observed, but these changes did not reach statistical significance. The SAR group with motivational feedback showed a slight decline in APSA scores post-intervention, whereas the tablet group without motivational feedback exhibited a minor improvement. This pattern could reflect differences in the cognitive demands or motivational aspects of the interventions. SARs, such as "Pepper," aim to provide motivational feedback, which has been shown to enhance engagement and persistence in tasks (26). However, the absence of significant improvements may be attributed to the short duration of the intervention, as cognitive training often requires sustained and repeated exposure to produce measurable gains in attention and performance (162). Additionally, the inpatient setting may have contributed to the limited improvements observed. Patients with depression often experience profound cognitive deficits, including impaired attention and memory, which are not easily ameliorated within a brief timeframe (92,163).

Sex differences in APSA scores were evident, with females consistently reporting higher performance in prospective memory and maintaining focused attention compared to males. These findings align with prior research demonstrating sex-related variations in cognitive performance. Women often outperform men in tasks requiring verbal memory, attention to detail, and multitasking, which are key components of prospective memory (164,165).
First, women may exhibit higher levels of internalized symptom awareness, reflecting broader sex-related differences in emotional processing, rumination tendencies, and self-monitoring of cognitive states (166). Second, sex-specific neurobiological pathways, such as estrogen-

related modulation of hippocampal and prefrontal activity, have been implicated in differential vulnerability to cognitive symptoms in depression (166–168). These factors may increase sensitivity to perceived deficits in prospective memory and attentional control, which are closely tied to daily-life functioning.

The persistence of these sex differences at post-intervention (t_1) further indicates that the short two-session cognitive training intervention, regardless of delivery mode, did not sufficiently modify subjective attention or performance perceptions. This stability over time is consistent with the broader literature showing that self-reported cognitive symptoms tend to be relatively resistant to brief interventions and may require longer or more intensive cognitive remediation programs (23). The persistence of sex differences post-intervention suggests that the cognitive training exercises did not disproportionately benefit one sex over the other. However, it remains unclear whether these differences reflect actual cognitive disparities or variations in self-assessment tendencies. Men may underestimate their cognitive performance due to cultural norms discouraging the acknowledgment of weaknesses, while women may be more self-critical and detailed in their evaluations (169,170).

While attention and performance outcomes provide insight into how participants perceived their cognitive functioning during the intervention, it is equally important to understand how they experienced the tools themselves. The following chapter therefore examines technology acceptance and usability, exploring whether the delivery mode (SAR-based with motivational feedback or tablet-based without motivational feedback) shaped participants' perceptions in ways that might influence engagement and feasibility, which may influence the longer-term effectiveness of cognitive training interventions.

Notably, age did not significantly influence technology acceptance or usability ratings, indicating that the acceptance and usability advantage of the robot was consistent across age groups. This homogeneity may be explained by the clinical setting, in which participants are accustomed to guided therapeutic activities and structured interventions, potentially attenuating demographic variability in acceptance-related attitudes (171,172).

In contrast, the TUI results revealed a more differentiated pattern, with significant age-related differences across several subscales. These differences were particularly pronounced in that younger participants reported higher overall technology usage, greater interest, higher perceived user-friendliness, and increased curiosity. These findings are consistent with broader demographic trends, whereby younger adults typically demonstrate higher digital affinity,

stronger exploratory motivation, and greater confidence in engaging with novel technologies (173). Elevated curiosity and interest scores suggest that younger participants may be more receptive to experimental or innovative technologies, which has practical implications for the age-sensitive tailoring of technology-based interventions in clinical settings.

While acceptance and usability measures provide insight into how participants perceived and interacted with technology, it is equally important to determine whether these experiences translated into differences in motivational engagement.

Regarding current motivation, the QCM results indicate that older participants (≥ 31 years) reported significantly higher fear of failure at baseline than younger participants. This age-related difference suggests that older adults may experience the training situation as more threatening or evaluative, possibly due to heightened concerns about performance, self-efficacy, or perceived expectations in a clinical setting. From a theoretical perspective, fear of failure is closely linked to avoidance-oriented achievement motivation and negative outcome expectancy. In the context of acute inpatient treatment, older adults may be more conscious of the implications of cognitive difficulties for daily functioning, occupational roles, or long-term autonomy, thereby intensifying self-evaluative pressure in performance-related tasks (174–178).

Interestingly, older participants also reported higher challenge scores at baseline compared to younger participants. This pattern suggests a more complex motivational profile in which avoidance-oriented fear of failure coexists with an approach-oriented tendency to perceive the task as challenging rather than merely threatening. Challenge appraisals have been associated with viewing demands as surmountable and personally meaningful, often co-occurring with moderate levels of anxiety or concern (179). In this sample, older adults may therefore simultaneously feel more at risk of failing while also perceiving the training as a relevant opportunity to test and improve their abilities. Such a dual appraisal is not uncommon in clinical populations, where individuals may be highly invested in recovery but also acutely aware of their limitations (180).

At the post-intervention measurement, participants in the tablet group without motivational feedback exhibited significantly higher challenge scores than those in the SAR group with motivational feedback. This finding may reflect differences in how the two delivery modes structure the performance context. The tablet version likely resembles more familiar digital task environments and may thus be experienced as a more straightforward, performance-focused situation (153,154). In contrast, the SAR-based condition, with its motivational feedback and

social presence, might buffer perceived difficulty and reduce the sense of being tested, thereby lowering challenge appraisals. While this could be beneficial in terms of reducing performance pressure and potential stress, it may also attenuate the perception of the task as a demanding achievement situation (179).

Beyond these specific findings, no significant group, sex, or age effects emerged for the QCM subscales probability of success and interest. This suggests that, across the sample, participants perceived comparable chances of performing well and reported similar levels of intrinsic interest in the task, irrespective of intervention modality or demographic characteristics. In other words, while older adults and tablet users differed in how challenging or threatening, they perceived the situation, their basic expectations of success and their interest in the activity remained relatively stable.

The QAM results provide complementary insights into more enduring achievement-related emotions and dispositions. At post-intervention, participants in the tablet group without motivational feedback reported significantly higher pride scores than those in the SAR group with motivational feedback. Pride in achievement contexts is typically associated with perceived success, competence, and the fulfillment of performance goals (181). The higher pride scores in the tablet condition may be interpreted as reflecting a stronger sense of having mastered a demanding cognitive task. This interpretation is consistent with the higher challenge ratings in the tablet group: when a situation is appraised as challenging and subsequently managed successfully, pride is a plausible outcome. In contrast, the SAR group's lower pride scores may again be linked to the SAR's more supportive and less evaluative interaction style, which could reduce both perceived challenge and the salience of personal achievement (26).

Sex differences emerged for shame at post-intervention, with female participants reporting significantly higher shame scores than male participants. Shame is a self-conscious emotion that reflects negative self-evaluation in the face of perceived failure or inadequacy (182,183). Higher shame scores among females are consistent with prior work suggesting that females may, on average, be more prone to internalizing negative feedback and attributing difficulties to personal deficiencies, particularly in achievement and performance contexts (184,185). In the setting of cognitive training for depression, this finding underscores the importance of gender-sensitive support strategies that minimize self-critical interpretations of performance and emphasize mastery, effort, and incremental progress (185).

No significant differences were found for confidence across group, sex, or age, although descriptive trends in some subgroups may suggest clinically meaningful variation that did not

reach statistical significance in this sample. The overall absence of broad effects on confidence indicates that the short intervention period was not sufficient to substantially alter participants' general achievement-related self-beliefs. This is consistent with theoretical models and empirical findings suggesting that confidence and self-efficacy are relatively stable constructs that change gradually over longer periods and repeated experiences (186).

While the quantitative motivational findings reveal these structured differences, they offer limited insight into how participants themselves made sense of the training experience, the role of the SAR or tablet in shaping their engagement, and the subjective meaning of challenge, pride, or shame in the context of their depressive symptoms. The following chapter therefore turns to the qualitative data to explore participants' perspectives in more detail and to complement the quantitative results with a deeper understanding of their lived experiences during the intervention

The general impressions of SAR "Pepper" highlight an initial ambivalence that gradually shifted toward greater comfort and familiarity. Participants described SAR "Pepper" as friendly, approachable, and non-threatening, reflecting design features intended to reduce anxiety and promote engagement. The fact that SAR "Pepper" was rarely perceived as gendered and instead described as "neutral" underscores the effectiveness of this anthropomorphic yet non-polarizing design. Such neutrality may be particularly advantageous in clinical mental health settings, where minimizing gender- or identity-related projections can support a more universally acceptable therapeutic presence. The reported initial unease, which diminished with continued interaction, aligns with previous research showing that users often require a brief acclimation period when interacting with SARs before positive engagement emerges (44).

SAR "Pepper's" voice emerged as a central point of criticism. Participants described the voice as artificial, monotonous, and insufficiently emotive. These perceptions are consistent with broader challenges in human–robot interaction research, where synthesized voices often struggle to convey the prosody, warmth, and variability necessary to support relational authenticity (26,187). The desire for customizable voice profiles points to the importance of adapting vocal characteristics to user preferences, which may enhance personalization and increase the perceived empathic capacity of the robot.

Reactions to SAR "Pepper's" movements and gestures further illustrate the importance of fine-tuned nonverbal communication. Smooth gestures were appreciated, yet sudden or repetitive movements caused discomfort or were perceived as malfunction-like. Participants' interpretations of these movements, sometimes attributing "nervousness" or emotional states to the robot, demonstrate the human tendency to ascribe agency and affect to embodied

systems. While this can enhance engagement when carefully calibrated, inconsistent or unexpected gestures may undermine trust and create distraction, echoing findings that motor expressiveness must be closely synchronized with verbal output to avoid cognitive dissonance (76).

A key finding concerns the perceived role of the SAR “Pepper” relative to the tablet interface. Although SAR “Pepper” was designed to deliver motivational feedback, participants reported that their primary interaction occurred through the tablet mounted on SAR “Pepper’s” chest. SAR “Pepper” functioned more as an auxiliary presence than as the main medium of training, a pattern that helps explain why TAM results did not show significant differences in perceived usefulness or intention to use: the tablet, rather than the robot, was the main operational interface. Technical interruptions, including loading delays and sporadic malfunctions, were particularly detrimental, disrupting concentration and diminishing trust in the system. This aligns with literature emphasizing that reliability is one of the strongest determinants of sustained engagement and acceptance in therapeutic technologies (26).

Motivationally, while SAR “Pepper’s” feedback was intended to enhance engagement, participants frequently critiqued the timing, contextual fit, and repetitiveness of the motivational statements. Encouragement that did not align with task performance, such as praise after incorrect responses, was perceived as disruptive and sometimes counterproductive. This observation reinforces the quantitative finding that motivational outcomes did not consistently favor the SAR condition. It also resonates with theories of motivational relevance, which emphasize that feedback must be contingent, timely, and performance-sensitive to foster competence and intrinsic motivation (188,189). The desire for adaptive, performance-linked feedback underscores the potential of future system enhancements using real-time performance monitoring, machine learning, or error-sensitive prompts.

Participants’ reflections on the cognitive training tasks themselves suggest that the exercises were perceived as meaningful and appropriately challenging, though sometimes overwhelming during acute depressive episodes. The absence of immediate error explanations and the inability to repeat tasks were common frustrations. These preferences highlight the need for more flexible training architecture that accounts for fluctuating cognitive load and daily mental state – a consideration particularly relevant for psychiatric inpatients, where cognitive performance can vary substantially from day to day.

Finally, participants’ views on trust and future use point toward cautious optimism. Many expressed openness to using SAR “Pepper” in broader therapeutic contexts, including outpatient or home-based training, provided that the system becomes more reliable and offers greater personalization. However, consistent with ethical debates in the field of SARs,

participants emphasized concerns related to data protection, autonomy, and the role of human therapists. The robot was viewed as a helpful addition but not a replacement for human care—a perspective well aligned with the conceptualization of SAR systems as supportive rather than substitutive tools in mental health care.

The qualitative insights shed light on the experiential and relational dimensions of interacting with SAR “Pepper,” complementing the quantitative results and identifying key areas for system improvement. To interpret these findings appropriately within the broader context of the study design and methodological approach, the following chapter outlines the strengths and limitations of the present investigation.

4.1 Strengths and Limitations

This study has several strengths and limitations that warrant discussion to contextualize its findings and implications. The methodological rigor and innovative approach contribute to the study's value, while certain limitations highlight areas for refinement in future research.

4.1.1 Strengths

A key strength of this study is the randomized group assignment, which effectively minimizes selection bias and ensures comparability between the intervention groups. Despite the lack of blinding, due to the visible nature of the interventions, the randomization process was rigorously implemented. This process was stratified by critical variables, including severity of depression, sex, and age categories, which effectively ensured balanced study groups. Furthermore, employing standardized procedures and consistent data collection protocols actively minimized potential sources of bias, thereby significantly enhancing the credibility of the findings (190–192).

Furthermore, the use of validated instruments (like the BDI-II, IDAS-II, APSA, TAM, SUS, TUI, and QCM) provides robust clinical assessments, significantly increasing the reliability of the findings. Another strength lies in employing multiple standardized measures to assess critical variables such as depressive symptoms, technology usability and acceptance, and motivation, allowing for a comprehensive evaluation.

The study's design further reflects its strength through the implementation of brief, structured interventions. This approach significantly reduced participant burden, a common challenge in clinical research. The streamlined methodology likely facilitated compliance and ensured consistent data quality, all while providing an accessible framework for testing innovative cognitive training interventions (147).

This pilot study advances research on SARs in mental health care by examining their clinical effects on depressive symptoms, attention and performance and motivation as well as on acceptance and usability of the SAR “Pepper” in psychiatric clinical setting. While previous research has primarily focused on older adults or individuals with cognitive impairments (82,193), this pilot study specifically addresses depression in a general adult population, thereby expanding the applicability of SAR interventions to broader psychiatric settings.

In addition to the quantitative outcome measures, incorporating qualitative data provided a holistic perspective on implementing the SAR "Pepper" in an acute psychiatric inpatient setting. While standardized instruments captured changes in depressive symptoms, cognitive functioning, usability, and motivational constructs, the qualitative findings illuminated patients' subjective experiences of the intervention, including emotional reactions, perceived usability, and relational aspects of human-robot interaction. This mixed-methods approach enabled a more comprehensive understanding of not only whether effects occurred, but also how and why participants responded to the technology in specific ways. By integrating objective outcome measures with personal narratives, the study offers a multidimensional evaluation of SAR "Pepper" that extends beyond symptom reduction to encompass the experiential, contextual, and practical dimensions of implementation in clinical mental health care.

4.1.2 Limitations

Despite its promising findings and strengths, this study has several limitations that warrant acknowledgment. The primary limitation concerns the asymmetry in feedback between the two groups. The SAR group received verbal motivational feedback from the SAR "Pepper", whereas the tablet condition provided no motivational feedback beyond task progression. This introduces an extraneous factor that may have contributed to the higher usability, acceptance, and motivation ratings observed in the SAR group. Consequently, the results should be interpreted as reflecting the combined effect of humanoid interaction and motivational feedback, rather than the robotic embodiment alone. Future studies must equalize feedback across conditions to isolate the specific contribution of the SAR component.

The second limitation was the relatively small sample size ($N = 32$), which may restrict the generalizability of the findings. Although the randomized design strengthens internal validity, the study’s statistical power remains limited, especially for detecting subtle or interaction effects

between variables. Future research should aim to replicate these findings in larger, more diverse samples to enhance external validity (148).

Another limitation relates to the brief intervention duration and low training frequency of the cognitive training. Participants completed only two training sessions, which may have been insufficient to produce measurable improvements in cognitive performance. Prior research indicates that long-term engagement is crucial for sustained benefits from SAR-based interventions, as engagement levels often decline after an initial novelty phase (67,162). Future research should explore the effects of prolonged SAR interventions on cognitive and emotional outcomes.

The sex- and age-specific differences observed in this study require further investigation. Intersectional aspects should also be considered to develop a nuanced understanding of the factors influencing SAR effectiveness in inpatients with depression. Eyssel and Hegel (194) emphasized the need to incorporate cultural and socioeconomic factors for a holistic perspective. Furthermore, this study focused exclusively on depression; future research should explore SAR use for other psychiatric conditions, such as anxiety disorders, bipolar disorders, and post-traumatic stress disorders. Studies by Huijnen et al. (195) and Rabbitt et al. (196) have already shown promising results for SAR use among patients with autism spectrum disorders and other psychiatric diseases.

The occurrence of technical problems with the SAR "Pepper" and the "MMA" application during the training sessions must also be mentioned. Although these issues were resolved during the interventions, such challenges may have negatively influenced participants' trust in this technology. Initial difficulties engaging with technology can skew users' perceptions of reliability and usability (26).

Another important limitation of the present study relates to baseline differences in depressive symptom severity between the intervention groups. Participants in the SAR group with motivational feedback exhibited significantly higher baseline BDI-II scores compared to those in the tablet group without motivational feedback. Although this imbalance was statistically controlled for in the analyses, such baseline differences may nonetheless affect the interpretation of intervention effects. Higher baseline severity of depressive symptoms is associated with increased heterogeneity in treatment response and may influence the magnitude and variability of observed changes over time. Previous research indicates that individuals with more severe depressive symptomatology can respond differently to

interventions than those with milder symptoms, potentially due to differences in cognitive functioning, emotional regulation, and physiological stress responses that shape treatment engagement and receptivity (148,150,155). Consequently, the observed effects should be interpreted with caution, as baseline symptom severity may have partially contributed to differential outcome trajectories between groups.

A further limitation of the present study is the absence of standardized measures assessing participants' prior knowledge and familiarity with technology. Although general technology acceptance and usage were assessed post-intervention, no standardized questionnaire was administered to evaluate participants' baseline level of technological knowledge before study entry. Pre-existing differences in technological knowledge may have influenced participants' engagement, perceived usability, and interaction with the intervention, independent of the actual effects of the SAR-based or tablet-based training (197). As this variable was neither systematically measured nor controlled for, its potential impact on the study outcomes cannot be fully ruled out and should be considered when interpreting the results (198).

Beyond these methodological and conceptual limitations, additional aspects warrant attention. First is the integration of SARs into existing treatment frameworks. While this study evaluated usability and acceptance, it did not examine how SAR-based cognitive training interacts with standard therapeutic approaches. Future studies should investigate whether combining SAR interventions with conventional treatments can enhance therapeutic outcomes, as such integrative approaches are crucial for realizing the full potential of SARs in psychiatric care (199). Second is the ethical dimension of SAR use in sensitive clinical environments. Although not the focus here, issues such as data privacy, informed consent, and unintended psychosocial consequences are highly relevant, especially with vulnerable populations. Addressing these ethical challenges proactively is essential for responsible implementation in healthcare robotics (200).

A further limitation concerns potential participant preferences and motivation toward the intervention modality. Some patients may have had a strong preference for training with the SAR "Pepper", while others may have been skeptical or less motivated toward robot-based interventions from the outset. Such pre-existing attitudes could have influenced engagement, adherence, and outcome measures independently of the actual intervention effects, thereby introducing a potential preference or expectancy bias. Although participants were allocated to either SAR-based or tablet-based cognitive training, individual motivation and acceptance may have differed substantially between groups.

4.2 Practical implications

The findings of this pilot study have several practical implications for the design and use of SAR-based cognitive training interventions for individuals with depression. Socially assistive robots such as “Pepper” offer specific advantages in clinical settings. In particular, their ability to provide motivational feedback and to engage users through interactive features may support cognitive training in people with depression. At the same time, the implementation of SARs in clinical practice requires careful planning. Potential challenges include technical reliability, users’ familiarity with the technology, and the additional resources needed for successful integration.

The significantly higher baseline severity of depressive symptoms in the SAR group, as measured by the BDI-II, highlights the importance of considering initial symptom burden when applying SAR-based interventions. Patients with more severe depressive symptoms may be more likely to be assigned to innovative or technology-supported treatments, either due to clinical decisions or patient preferences. From a practical perspective, this underlines the need for careful stratification or matching procedures in future implementations in order to avoid systematic baseline differences between groups. Clinically, the results also suggest that SAR-based cognitive training can be feasible even for individuals with more severe depression. However, expectations regarding short-term symptom improvement should be adjusted accordingly.

Sex-related differences were most evident in attention and performance outcomes. Female participants consistently reported higher overall attention and performance, as well as better prospective memory and focused attention, both at baseline and after the intervention. These stable differences suggest that female patients may experience or perceive higher cognitive functioning during inpatient treatment, regardless of the intervention type. In clinical practice, this indicates that male patients may benefit from additional guidance, structure, or motivational support during cognitive training, especially when interventions are delivered through digital or semi-autonomous systems. SAR-based interventions may therefore benefit from sex-sensitive design features, such as adaptive pacing, clearer performance feedback, or additional prompts to help maintain attention in male users.

Age-related results further emphasize the importance of tailoring SAR-based interventions to different user groups. Although no age differences were found in depressive symptoms,

attention performance, or general technology acceptance and usability, younger participants reported higher overall technology use, greater interest, higher perceived user-friendliness, and greater curiosity. These findings suggest that younger adults may be more open to interacting with SARs and digital technologies in general, even when clinical outcomes do not differ by age. In practice, SAR-based cognitive training may therefore be particularly suitable as a low-threshold or introductory digital intervention for younger patients. Older patients, in contrast, may require more extensive introduction, reassurance, and ongoing technical support in order to reach similar levels of engagement and acceptance.

Motivational findings provide further guidance for practical implementation. Older participants reported higher fear of failure and higher perceived challenge at baseline, indicating a more performance-sensitive motivational profile. This suggests that SAR-based interventions for older adults should reduce evaluative pressure and focus on supportive, non-judgmental feedback. At the same time, participants in the tablet group reported higher perceived challenge and greater achievement-related pride after the intervention. This indicates that self-directed digital formats may foster a stronger sense of accomplishment for some users. From a practical perspective, these findings highlight the risk that poorly timed or non-adaptive motivational feedback from a SAR may reduce feelings of autonomy or competence. This underlines the importance of carefully designed and well-timed feedback strategies.

The qualitative findings complement the quantitative results and offer important insights for intervention design. Participants mainly perceived SAR “Pepper” as a supportive companion rather than as the primary medium of cognitive training. The tablet interface remained the central element of interaction. This suggests that, in clinical practice, SARs may be most effective when used as motivational or structuring supports rather than as standalone training platforms. Technical reliability was identified as a key factor for acceptance, as even short technical problems or delays were reported to reduce trust and concentration. Reliable system performance and minimal delays are therefore essential for real-world use.

Participants also clearly emphasized the need for personalized and adaptive motivational feedback. Standardized and repetitive encouragement was often experienced as distracting, especially when it did not match actual task performance. From a practical standpoint, SAR-based interventions should therefore include feedback mechanisms that adapt to individual performance, effort, and progress. Such personalization may be particularly important for older users and for individuals with higher fear of failure.

Overall, the findings indicate that SAR-based cognitive training interventions should not be implemented as uniform, one-size-fits-all solutions. Instead, practical implementation should take into account baseline symptom severity, sex-related differences in attention and self-evaluation, age-related differences in technology engagement, and individual motivational profiles. When these factors are systematically considered, SARs such as “Pepper” may serve as valuable additions to established therapeutic approaches, particularly by improving engagement and providing structure in digitally supported cognitive training for individuals with depression.

4.3 Future research directions

Building on the findings and limitations of this pilot study, several directions for future research can be identified that are essential for advancing SAR-based interventions in psychiatric care. These directions mainly relate to study design, the investigation of underlying mechanisms, individual differences between users, and the integration of SARs into clinical practice.

A key priority for future research is to clearly separate the effects of robotic embodiment from those of motivational feedback. In the present study, the SAR condition combined a humanoid robot with verbal motivational feedback, whereas the tablet condition did not provide comparable feedback. As a result, it remains unclear whether the observed differences in usability, acceptance, and motivation are attributable to the robot itself or to the motivational elements. Future studies should therefore use for example a multi-arm randomized controlled designs that independently vary embodiment (e.g., humanoid robot vs. non-humanoid system vs. tablet-based intervention) and feedback type (e.g., motivational vs. neutral feedback). Such designs would allow a more precise examination of how different components contribute to engagement and outcomes.

In addition, future research should focus more closely on the design of motivational feedback. Qualitative findings from this study suggest that standardized and non-adaptive feedback can be perceived as repetitive or distracting. Future studies should therefore investigate adaptive feedback systems that respond to task performance, effort, or changes in cognitive or emotional state. Comparing static and adaptive feedback, as well as different feedback styles (e.g., supportive or informational), may help identify which approaches best support sustained engagement without reducing perceived autonomy. Advances in personalization, including algorithm-based or learning systems, may further enhance the effectiveness of SAR-based interventions.

The observed sex- and age-related differences underline the importance of considering individual characteristics in future research. Studies with larger sample sizes are needed to examine whether differences in attention, motivation, and technology-related attitudes lead to different intervention effects over time. Beyond sex and age, factors such as prior experience with technology, baseline cognitive functioning, symptom severity, and psychosocial variables should be included. This would allow a more differentiated understanding of which patient groups benefit most from specific intervention designs.

Another important direction concerns the duration and intensity of SAR-based interventions. In this study, the short intervention period and low training frequency limit conclusions about cognitive and emotional change. Future research should therefore use longer intervention periods and include follow-up assessments to evaluate long-term effects, adherence, and possible novelty effects. Longitudinal study designs can capture individual change trajectories and identify optimal intervention schedules.

From a methodological perspective, cross-over may be particularly useful in future studies. Allowing participants to experience different intervention modalities in a randomized order can reduce the influence of individual preferences, expectations, and baseline motivation. Such designs may provide more detailed insights into acceptance, usability, and perceived effectiveness, especially in heterogeneous psychiatric inpatient populations.

Future research should systematically assess participants' baseline level of technological knowledge using standardized and validated questionnaires. Such assessments should be conducted prior to randomization in order to capture pre-existing differences in familiarity with digital and robotic technologies. Beyond being included as control variables in statistical analyses, baseline technology knowledge could be incorporated as an additional stratification factor during randomization. This approach would help ensure a balanced distribution of technological knowledge across intervention groups and reduce potential confounding effects. Incorporating standardized measures of technological knowledge may therefore improve internal validity and allow for a more precise interpretation of the specific effects of SAR-based interventions.

Future research should also extend beyond depression and examine SAR-based interventions across a broader range of psychiatric conditions, such as anxiety disorders, bipolar disorders, post-traumatic stress disorder, or other mental health conditions. Comparative studies across

diagnostic groups may help identify shared mechanisms, such as attention regulation or motivation, that are especially relevant for SAR-supported interventions.

Finally, future studies should increasingly examine how SAR-based interventions can be integrated into existing treatment settings. Rather than focusing on SARs as standalone tools, research should explore their use as complementary elements within established therapeutic approaches. At the same time, ethical and organizational aspects, including data protection, and the role of human therapists, should be considered to support responsible implementation in clinical care.

5. Conclusion

Across outcome domains, the intervention did not yield differential effects on depressive symptom severity or self-reported cognitive functioning between the SAR-based and tablet-based conditions within the short intervention period. These findings suggest that brief cognitive training, irrespective of delivery modality, may be insufficient to induce measurable changes in core depressive symptoms or subjective attention and performance outcomes in acutely ill inpatient populations. However, substantial sex differences in perceived cognitive difficulties and motivational responses highlight the importance of gender-sensitive interpretation when evaluating cognitive and emotional self-assessments in depression.

In contrast to the largely comparable clinical outcomes, the SAR-based intervention with motivational feedback produced significantly higher usability ratings and was perceived as more user-friendly, indicating an experiential advantage of embodied technology over conventional tablet delivery. At the same time, the absence of group differences in technology acceptance suggests that short-term exposure may not be sufficient to modify perceived usefulness or behavioral intention, underscoring the conceptual distinction between usability and acceptance.

The qualitative findings further revealed that while users responded positively to the SAR “Pepper’s” presence and novelty, limitations in voice naturalness, gesture coherence, and technical reliability constrained deeper emotional engagement and trust.

Motivational outcomes revealed a differentiated pattern shaped by age, sex, and intervention modality rather than a uniform effect. Older participants demonstrated heightened fear of failure alongside stronger challenge appraisals, female participants reported higher shame, and tablet users experienced greater pride and challenge. These results indicate that cognitive training environments may differentially activate motivational and self-evaluative processes across demographic subgroups, with potentially relevant implications for intervention personalization.

6. References

1. World Health Organization (WHO). Depressive disorder (depression). Depressive disorder (depression). 2023.

2. Nowotny M, Kern D, Breyer E, Bengough T, Griebler R. Depressionsbericht Österreich. Eine interdisziplinäre und multiperspektivische Bestandsaufnahme [Internet]. Wien: Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz; 2019. Available from: https://www.sozialministerium.at/Themen/Gesundheit/Nicht-uebertragbare-Krankheiten/Psychische-Gesundheit/Depressionsbericht-%C3%96sterreich.html

3. Zhu C, Zhang T, Li Q, Chen X, Wang K. Depression and Anxiety During the COVID-19 Pandemic: Epidemiology, Mechanism, and Treatment. Neurosci Bull. 2023 Apr;39(4):675–84.

4. Cortés Zamora EB, Mas Romero M, Tabernero Sahuquillo MT, Avendaño Céspedes A, Andrés-Petrel F, Gómez Ballesteros C, et al. Psychological and Functional Impact of COVID-19 in Long-Term Care Facilities: The COVID-A Study. The American Journal of Geriatric Psychiatry. 2022 Apr;30(4):431–43.

5. American Psychiatric Association. What Is Depression? [Internet]. 2026 [cited 2026 Jan 30]. Available from: https://www.psychiatry.org:443/patients-families/depression/what-is-depression

6. Rothenhäusler HB, Täschner KL. Kompendium Praktische Psychiatrie: und Psychotherapie. 2. Aufl. 2012. Vienna: Springer Vienna Imprint Springer; 2012.

7. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and Acceptability of Cognitive Behavior Therapy Delivery Formats in Adults With Depression: A Network Meta-analysis. JAMA Psychiatry. 2019 Jul 1;76(7):700.

8. Malhi GS, Bell E, Bassett D, Boyce P, Bryant R, Hazell P, et al. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. Aust N Z J Psychiatry. 2021 Jan;55(1):7–117.

9. Stubbe DE. Abating Ambivalence: Motivating Depressed Patients for Treatment. Focus (Am Psychiatr Publ). 2020 Apr;18(2):197–200.

10. Russo M, Mahon K, Burdick KE. MEASURING COGNITIVE FUNCTION IN MDD: EMERGING ASSESSMENT TOOLS. Depress Anxiety. 2015 Apr;32(4):262–9.

11. Ahern E, Semkovska M. Cognitive functioning in the first-episode of major depressive disorder: A systematic review and meta-analysis. Neuropsychology. 2017 Jan;31(1):52–72.

12. Kaser M, Zaman R, Sahakian BJ. Cognition as a treatment target in depression. Psychol Med. 2017 Apr;47(6):987–9.

13. Konrad A, Bellotti V, Crenshaw N, Tucker S, Nelson L, Du H, et al. Finding the Adaptive Sweet Spot: Balancing Compliance and Achievement in Automated Stress Reduction. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing

Systems [Internet]. Seoul Republic of Korea: ACM; 2015 [cited 2023 Jun 20]. p. 3829–38. Available from: https://dl.acm.org/doi/10.1145/2702123.2702512

14. Gkintoni E, Vassilopoulos SP, Nikolaou G. Mindfulness-Based Cognitive Therapy in Clinical Practice: A Systematic Review of Neurocognitive Outcomes and Applications for Mental Health and Well-Being. JCM. 2025 Mar 3;14(5):1703.

15. Nakao M, Shirotsuki K, Sugaya N. Cognitive-behavioral therapy for management of mental health and stress-related disorders: Recent advances in techniques and technologies. Biopsychosoc Med. 2021 Oct 3;15(1):16.

16. Diamond A. Executive Functions. Annu Rev Psychol. 2013 Jan 3;64(1):135–68.

17. Gotlib IH, Joormann J. Cognition and Depression: Current Status and Future Directions. Annu Rev Clin Psychol. 2010 Mar 1;6(1):285–312.

18. Sachs-Ericsson N, Joiner T, Plant EA, Blazer DG. The influence of depression on cognitive decline in community-dwelling elderly persons. The American journal of geriatric psychiatry. 2005;13(5):402–8.

19. Scheurich A, Fellgiebel A, Schermuly I, Bauer S, Wölfges R, Müller MJ. Experimental evidence for a motivational origin of cognitive impairment in major depression. Psychol Med. 2008 Feb;38(2):237–46.

20. Wood-Ross C, Tran T, Milanovic M, Jokic R, Milev R, Bowie CR. Neurocognition and Depressive Symptoms have Unique Pathways to Predicting Different Domains of Functioning in Major Depressive Disorder. Can J Psychiatry. 2023 Apr;68(4):241–8.

21. Akbari Farkhani V. Computerized Cognitive Training Interventions in Major Depression. Iran J Psychiatry Behav Sci [Internet]. 2022 Mar 19 [cited 2025 Dec 23];16(1). Available from: https://brieflands.com/journals/ijpbs/articles/121464

22. Bursky M, Egglefield DA, Schiff SG, Premnath P, Sneed JR. Mindfulness-Enhanced Computerized Cognitive Training for Depression: An Integrative Review and Proposed Model Targeting the Cognitive Control and Default-Mode Networks. Brain Sciences. 2022 May 19;12(5):663.

23. Motter JN, Pimontel MA, Rindskopf D, Devanand DP, Doraiswamy PM, Sneed JR. Computerized cognitive training and functional recovery in major depressive disorder: A meta-analysis. Journal of Affective Disorders. 2016 Jan;189:184–91.

24. Sala G, Gobet F. Cognitive Training Does Not Enhance General Cognition. Trends in Cognitive Sciences. 2019 Jan;23(1):9–20.

25. Fortuna A, Prasetya F, Samala AD, Rawas S, Criollo-C S, Kaya D, et al. Artificial intelligence in personalized learning: A global systematic review of current advancements and shaping future opportunities. Social Sciences & Humanities Open. 2025;12:102114.

26. Broadbent E. Interactions With Robots: The Truths We Reveal About Ourselves. Annu Rev Psychol. 2017 Jan 3;68(1):627–52.

27. Chen S, Jones C, Moyle W. Social Robots for Depression in Older Adults: A Systematic Review. Journal of Nursing Scholarship. 2018 Nov;50(6):612–22.

28. Feil-Seifer D, Matarić MJ. Defining socially assistive robotics. 9th International Conference on Rehabilitation Robotics, 2005 ICORR 2005. 2005;465–8.

29. Mollahosseini A, Abdollahi H, Sweeny TD, Cole R, Mahoor MH. Role of embodiment and presence in human perception of robots' facial cues. International Journal of Human-Computer Studies. 2018 Aug;116:25–39.

30. Raptopoulou A, Komnidis A, Bamidis PD, Astaras A. Human–robot interaction for social skill development in children with ASD: A literature review. Healthcare Tech Letters. 2021 Aug;8(4):90–6.

31. Ahmed MM, Okesanya OJ, Olaleke NO, Adigun OA, Adebayo UO, Oso TA, et al. Integrating Digital Health Innovations to Achieve Universal Health Coverage: Promoting Health Outcomes and Quality Through Global Public Health Equity. Healthcare (Basel). 2025 May 5;13(9):1060.

32. Reininghaus E, Dalkner N. Robotics in psychiatry – Fiction or reality? European Neuropsychopharmacology. 2023 May;70:17–8.

33. Breazeal C. Toward sociable robots. Robotics and Autonomous Systems. 2003 Mar;42(3–4):167–75.

34. Breazeal C. Emotion and sociable humanoid robots. International Journal of Human-Computer Studies. 2003 Jul;59(1–2):119–55.

35. Weng YH, Chen CH, Sun CT. Toward the Human–Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots. Int J of Soc Robotics. 2009 Nov;1(4):267–82.

36. Hurrell A. The retirement homes of the future are being planned today. 2017;

37. Lorenz T, Weiss A, Hirche S. Synchrony and Reciprocity: Key Mechanisms for Social Companion Robots in Therapy and Care. Int J of Soc Robotics. 2016 Jan;8(1):125–43.

38. Jack RE. Teaching robots the art of human social synchrony. Sci Robot. 2024 Mar 27;9(88):eado5755.

39. Niculescu AI, Yeo KH, Ehnes J. Am I a Social Buddy? A Literature Review on Socially Appealing Design and Implementation Methods for Social Robots. In: Li H, Schultz T, Bi Y, Zhu J, He H, Ma J, et al., editors. Social Robotics [Internet]. Singapore: Springer Nature Singapore; 2025 [cited 2025 Nov 12]. p. 187–96. (Lecture Notes in Computer Science; vol. 15170). Available from: https://link.springer.com/10.1007/978-981-96-1151-5_19

40. Giorgi I, Minutolo A, Tirotto F, Hagen O, Esposito M, Gianni M, et al. I am Robot, Your Health Adviser for Older Adults: Do You Trust My Advice? Int J of Soc Robotics. 2025 May;17(5):883–902.

41. Welsch R, Von Castell C, Hecht H. The anisotropy of personal space. Wilson RK, editor. PLoS ONE. 2019 Jun 4;14(6):e0217587.

42. Prakash A, Kemp CC, Rogers WA. Older adults' reactions to a robot's appearance in the context of home use. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction [Internet]. Bielefeld Germany: ACM; 2014 [cited 2023 Jun 20]. p. 268–9. Available from: https://dl.acm.org/doi/10.1145/2559636.2559790

43. Laakasuo M, Palomäki J, Köbis N. Moral Uncanny Valley: A Robot's Appearance Moderates How its Decisions are Judged. Int J of Soc Robotics. 2021 Nov;13(7):1679–88.

44. de Graaf MMA, Allouch SB, Klamer T. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. Computers in Human Behavior. 2015 Feb;43:1–14.

45. Payr S. Towards Human–Robot Interaction Ethics. In: Trappl R, editor. A Construction Manual for Robots' Ethical Systems [Internet]. Cham: Springer International Publishing; 2015 [cited 2023 Jun 20]. p. 31–62. (Cognitive Technologies). Available from: https://link.springer.com/10.1007/978-3-319-21548-8_3

46. Zuschnegg J, Häussl A, Lodron G, Orgel T, Russegger S, Schneeberger M, et al. Psychosocial effects of a humanoid robot on informal caregivers of people with dementia: A randomised controlled trial with nested interviews. International Journal of Nursing Studies. 2025 Feb;162:104967.

47. Causo A, Vo GT, Chen IM, Yeo SH. Design of Robots Used as Education Companion and Tutor. In: Zeghloul S, Laribi MA, Gazeau JP, editors. Robotics and Mechatronics [Internet]. Cham: Springer International Publishing; 2016 [cited 2023 Feb 28]. p. 75–84. (Mechanisms and Machine Science; vol. 37). Available from: https://link.springer.com/10.1007/978-3-319-22368-1_8

48. Henkel Z, Henkel KB, Thompson J, Smith K, Stubbs-Richardson M, Bethel CL. The Socially Therapeutic Assistive Robot (STAR) Therabot and Its Modular Sensor Collar. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction [Internet]. Boulder CO USA: ACM; 2024 [cited 2025 Mar 19]. p. 77–9. Available from: https://dl.acm.org/doi/10.1145/3610978.3641088

49. Szondy M, Fazekas P. Attachment to robots and therapeutic efficiency in mental health. Front Psychol. 2024 Jan 31;15:1347177.

50. Zuschnegg J, Paletta L, Fellner M, Steiner J, Pansy-Resch S, Jos A, et al. Humanoid socially assistive robots in dementia care: a qualitative study about expectations of caregivers and dementia trainers. Aging & Mental Health. 2022 Jun 3;26(6):1270–80.

51. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 1989 Sep;13(3):319.

52. Markgraf D. Akzeptanztheorie - Definition: Was ist "Akzeptanztheorie"? [Internet]. 2018. Available from: https://wirtschaftslexikon.gabler.de/definition/akzeptanztheorie-27148/version-250811 Revision von Akzeptanztheorie

53. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science. 2000 Feb;46(2):186–204.

54. Ajzen I, Fishbein M. A Bayesian analysis of attribution processes. Psychological Bulletin. 1975 Mar;82(2):261–77.

55. Rauniar R, Rawski G, Yang J, Johnson B. Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. Journal of Enterprise Information Management. 2014 Feb 4;27(1):6–30.

56. Habets W, Ramakers W, Hansen D, Coninx K. Technology acceptance of the EXPERT tool for guideline-based exercise prescription is strongly influenced by the organizational context. European Journal of Preventive Cardiology. 2024 Jun 13;31(Supplement_1):zwae175.205.

57. Klawunn R, Albrecht UV, Katzmarzyk D, Dierks ML. Beyond technology acceptance—a focused ethnography on the implementation, acceptance and use of new nursing technology in a German hospital. Front Digit Health. 2024 Apr 25;6:1330988.

58. Madi M, Nielsen S, Schweitzer M, Siebert M, Körner D, Langensiepen S, et al. Acceptance of a robotic system for nursing care: a cross-sectional survey with professional nurses, care recipients and relatives. BMC Nurs. 2024 Mar 14;23(1):179.

59. Shania F, Paramarta V. Analysis of Technology Acceptance Model (TAM) on The Use of Electronic Medical Records in Hospitals. JISS. 2024 Dec 25;5(12):3190–6.

60. Wicki S, Clark IC, Amann M, Christ SM, Schettle M, Hertler C, et al. Acceptance of Digital Health Technologies in Palliative Care Patients. Palliative Medicine Reports. 2024 Jan 1;5(1):34–42.

61. Aboujaoude E, Gega L, Parish MB, Hilty DM. Editorial: Digital Interventions in Mental Health: Current Status and Future Directions. Front Psychiatry. 2020;11:111.

62. Holden RJ, Karsh BT. The Technology Acceptance Model: Its past and its future in health care. Journal of Biomedical Informatics. 2010 Feb;43(1):159–72.

63. Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. Int J of Soc Robotics. 2009 Jan;1(1):71–81.

64. Mori M, MacDorman K, Kageki N. The Uncanny Valley [From the Field]. IEEE Robot Automat Mag. 2012 Jun;19(2):98–100.

65. Dinesen B, Hansen HK, Grønborg GB, Dyrvig AK, Leisted SD, Stenstrup H, et al. Use of a Social Robot (LOVOT) for Persons With Dementia: Exploratory Study. JMIR Rehabil Assist Technol. 2022 Aug 1;9(3):e36505.

66. Lee H, Chung MA, Kim H, Nam EW. The Effect of Cognitive Function Health Care Using Artificial Intelligence Robots for Older Adults: Systematic Review and Meta-analysis. JMIR Aging. 2022 Jun 28;5(2):e38896.

67. Piatt J, Nagata S, Šabanović S, Cheng WL, Bennett C, Lee HR, et al. Companionship with a robot? Therapists' perspectives on socially assistive robots as therapeutic interventions in community mental health for older adults. Am J of Rec Therapy. 2017 Feb 12;15(4):29.

68. Lee OEK, Nam I, Chon Y, Park A, Choi N. Socially Assistive Humanoid Robots: Effects on Depression and Health-Related Quality of Life among Low-Income, Socially Isolated Older Adults in South Korea. J Appl Gerontol. 2023 Mar;42(3):367–75.

69. Kling M, Haeussl A, Dalkner N, Fellendorf FT, Lenger M, Finner A, et al. Social robots in adult psychiatry: a summary of utilisation and impact. Front Psychiatry. 2025 Feb 11;17:1506776.

70. Nichol B, McCready J, Erfani G, Comparcini D, Simonetti V, Cicolini G, et al. Exploring the impact of socially assistive robots on health and wellbeing across the lifespan: An umbrella review and meta-analysis. International Journal of Nursing Studies. 2024 May;153:104730.

71. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially Assistive Robots in Elderly Care: A Systematic Review into Effects and Effectiveness. Journal of the American Medical Directors Association. 2012;13(2):114-120.e1.

72. Moyle W, Jones C, Murfield J, Thalib L, Beattie E, Shum D, et al. Effect of a robotic seal on the motor activity and sleep patterns of older people with dementia, as measured by wearable technology: A cluster-randomised controlled trial. Maturitas. 2018 Apr;110:10–7.

73. Otaka E, Osawa A, Kato K, Obayashi Y, Uehara S, Kamiya M, et al. Positive Emotional Responses to Socially Assistive Robots in People With Dementia: Pilot Study. JMIR Aging. 2024 Apr 11;7:e52443.

74. Xie J. The Application of Social Robots in Healthcare: Enhancing Care Quality and Addressing Ethics Challenges. ACE. 2024 Dec 12;114(1):128–32.

75. Zuschnegg J, Paletta L, Fellner M, Steiner J, Pansy-Resch S, Jos A, et al. Humanoid socially assistive robots in dementia care: a qualitative study about expectations of caregivers and dementia trainers. Aging & Mental Health. 2022;26(6):1270–80.

76. Zuschnegg J, Häussl A, Lodron G, Orgel T, Russegger S, Schneeberger M, et al. Psychosocial effects of a humanoid robot on informal caregivers of people with dementia: A randomised controlled trial with nested interviews. International Journal of Nursing Studies. 2025;162(null):N.PAG-N.PAG.

77. Winkle K, Caleb-Solly P, Turton A, Bremner P. 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 [Internet]. Chicago IL USA: ACM; 2018 [cited 2025 Jan 5]. p. 289–97. Available from: https://dl.acm.org/doi/10.1145/3171221.3171273

78. Mann JA, MacDonald BA, Kuo IH, Li X, Broadbent E. People respond better to robots than computer tablets delivering healthcare instructions. Computers in Human Behavior. 2015 Feb;43:112–7.

79. Chen SC, Moyle W, Jones C, Petsky H. A social robot intervention on depression, loneliness, and quality of life for Taiwanese older adults in long-term care. Int Psychogeriatr. 2020 Aug;32(8):981–91.

80. Lim J. Effects of a cognitive-based intervention program using social robot PIO on cognitive function, depression, loneliness, and quality of life of older adults living alone. Front Public Health. 2023 Feb 6;11:1097485.

81. Duradoni M, Colombini G, Russo PA, Guazzini A. Robotic Psychology: A PRISMA Systematic Review on Social-Robot-Based Interventions in Psychological Domains. J. 2021 Oct 26;4(4):664–709.

82. Dino F, Zandie R, Abdollahi H, Schoeder S, Mahoor MH. Delivering Cognitive Behavioral Therapy Using A Conversational Social Robot. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) [Internet]. Macau, China: IEEE; 2019 [cited 2025 Mar 3]. p. 2089–95. Available from: https://ieeexplore.ieee.org/document/8968576/

83. Scassellati B, Henny Admoni, Matarić M. Robots for Use in Autism Research. Annu Rev Biomed Eng. 2012 Aug 15;14(1):275–94.

84. Jung HW, Park JY, Holoubek T, Kim WJ, Park J. Socially Assistive Robots in Mental Healthcare: Principles and Conceptual Framework for User-Centered Design. Int J of Soc Robotics [Internet]. 2025 Sep 18 [cited 2025 Nov 11]; Available from: https://link.springer.com/10.1007/s12369-025-01323-5

85. Spitale M, Silleresi S, Garzotto F, Matarić MJ. Using Socially Assistive Robots in Speech-Language Therapy for Children with Language Impairments. Int J of Soc Robotics. 2023 Oct;15(9–10):1525–42.

86. Vagnetti R, Di Nuovo A, Mazza M, Valenti M. Social Robots: A Promising Tool to Support People with Autism. A Systematic Review of Recent Research and Critical Analysis from the Clinical Perspective. Rev J Autism Dev Disord [Internet]. 2024 Feb 29 [cited 2025 Nov 11]; Available from: https://link.springer.com/10.1007/s40489-024-00434-5

87. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Effectiveness of Robot Paro in Intramural Psychogeriatric Care: A Multicenter Quasi-Experimental Study. Journal of the American Medical Directors Association. 2015 Nov;16(11):946–50.

88. Blanson Henkemans OA, Bierman BPB, Janssen J, Neerincx MA, Looije R, van der Bosch H, et al. Using a robot to personalise health education for children with diabetes type 1: A pilot study. Patient Education and Counseling. 2013 Aug;92(2):174–81.

89. Costescu CA, Vanderborght B, David DO. The Effects of Robot-Enhanced Psychotherapy: A Meta-Analysis. Review of General Psychology. 2014 Jun;18(2):127–36.

90. Pennisi P, Tonacci A, Tartarisco G, Billeci L, Ruta L, Gangemi S, et al. Autism and social robotics: A systematic review. Autism Research. 2016 Feb;9(2):165–83.

91. World Health Organization (WHO). Mental health: strengthening our response. [Internet]. Mental health. 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

92. Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med. 2014 Jul;44(10):2029–40.

93. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of Social Robots in Mental Health and Well-Being Research: Systematic Review. J Med Internet Res. 2019 Jul 24;21(7):e13322.

94. Łukasik S, Tobis S, Suwalska J, Łojko D, Napierała M, Proch M, et al. The Role of Socially Assistive Robots in the Care of Older People: To Assist in Cognitive Training, to Remind or to Accompany? Sustainability. 2021 Sep 17;13(18):10394.

95. Weerarathna IN, Raymond D, Luharia A. Human-Robot Collaboration for Healthcare: A Narrative Review. Cureus. 2023 Nov;15(11):e49210.

96. Välimäki M, Hätönen H, Adams CE. Mobile.net: Mobile Telephone Text Messages to Encourage Adherence to Medication and to Follow up With People With Psychosis: Methods and Protocol for a Multicenter Randomized Controlled Two-Armed Trial. JMIR Res Protoc. 2012 Aug 2;1(2):e8.

97. Gefen, Karahanna, Straub. Trust and TAM in Online Shopping: An Integrated Model. MIS Quarterly. 2003;27(1):51.

98. Nordhues HC, Bhagra A, Stroud NN, Vencill JA, Kuhle CL. COVID-19 Gender Disparities and Mitigation Recommendations: A Narrative Review. Mayo Clin Proc. 2021 Jul;96(7):1907–20.

99. Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025 Apr 14;e081123.

100. Hengartner MP, Jakobsen JC, Sørensen A, Plöderl M. Efficacy of new-generation antidepressants assessed with the Montgomery-Asberg Depression Rating Scale, the gold standard clinician rating scale: A meta-analysis of randomised placebo-controlled trials. Hashimoto K, editor. PLoS ONE. 2020 Feb 26;15(2):e0229381.

101. Strauß B, Schumacher J. Klinische Interviews und Ratingskalen. 1. Auflage. Göttingen: Hogrefe Verlag; 2005.

102. Hautzinger M, Keller F, Kühner C. BDI-II. Beck-Depressions-Inventar. Revision. 2.Auflage. Frankfurt: Pearson Assessment; 2009.

103. Maher JM, Markey JC, Ebert-May D. The other half of the story: effect size analysis in quantitative research. CBE Life Sci Educ. 2013;12(3):345–51.

104. Pszeida M, Häussl A, Zuschnegg J, Macher M, Draxler S, Orgel T, et al. Motivating Patients with Depression for Gender-sensitive Cognitive Training Using a Socially Assistive Robot with Bio-signal Driven Pause Management. In 2025 [cited 2025 Nov 10]. Available from: https://openaccess.cms-conferences.org/publications/book/978-1-964867-41-0/article/978-1-964867-41-0_5

105. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods. 2009 Nov;41(4):1149–60.

106. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods. 2007 May;39(2):175–91.

107. Study Randomizer [Internet]. 2017; Available from: https://www.studyrandomizer.com

108. Wang X, Piantadosi S, Le-Rademacher J, Mandrekar SJ. Statistical Considerations for Subgroup Analyses. J Thorac Oncol. 2021 Mar;16(3):375–80.

109. Kang NE, Yoon WC. Age- and experience-related user behavior differences in the use of complicated electronic devices. International Journal of Human-Computer Studies. 2008 Jun;66(6):425–37.

110. Mehta CM, Arnett JJ, Palmer CG, Nelson LJ. Established adulthood: A new conception of ages 30 to 45. American Psychologist. 2020 May;75(4):431–44.

111. Mollborn S, Fomby P, Goode JA, Modile A. A life course framework for understanding digital technology use in the transition to adulthood. Adv Life Course Res. 2021 Mar;47:100379.

112. Haeussl A, Zwigl I, Stojec L, Smolak I, Stijic M, Stross T, et al. Evaluating usability and acceptance of a socially assistive robot supported cognitive training for depression – results of the randomized controlled pilot study 'AMIGA.' Front Psychiatry. 2026 Jan 15;16:1661730.

113. Aldebaran. Aldebaran. [cited 2025 Jun 7]. Aldebaran - Pepper. Available from: https://aldebaran.com/en/pepper/

114. Probo Robotics. Technische Spezifikationen. [Internet]. 2022 [cited 2022 Nov 22]. Available from: https://www.probo-robotics.at/pepper-daten-fakten/.

115. SoftBank Robotics Group Corp. SoftBank Robotics Group Corp. 2023 [cited 2025 Jun 7]. Solution. Available from: https://www.softbankrobotics.com/solution/

116. Paletta L, Fellner M, Pszeida M, Lerch A, Kemp C, Pittino L, et al. Playful Multimodal Training for Persons with Dementia with Executive Function based Decision Support. In: Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference [Internet]. Corfu Greece: ACM; 2018 [cited 2025 Jun 10]. p. 237–40. Available from: https://dl.acm.org/doi/10.1145/3197768.3197789

117. Paletta L, Brijacak I, Reiterer B, Pszeida M, Ganster H, Fuhrmann F, et al. Gaze-based Human Factors Measurements for the Evaluation of Intuitive Human-Robot Collaboration in Real-time. In: 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA) [Internet]. Zaragoza, Spain: IEEE; 2019 [cited 2023 Jun 20]. p. 1528–31. Available from: https://ieeexplore.ieee.org/document/8869270/

118. Paletta L, Pszeida M, Fellner M, Russegger S, Dini A, Draxler S, et al. Towards Decision Support with Assessment of Neuropsychological Profiles in Alzheimer's Dementia Using Playful Tablet-Based Multimodal Activation. In: Ayaz H, Asgher U, Paletta L, editors. Advances in Neuroergonomics and Cognitive Engineering [Internet]. Cham: Springer International Publishing; 2021 [cited 2025 Jun 10]. p. 469–77. (Lecture Notes in Networks and Systems; vol. 259). Available from: https://link.springer.com/10.1007/978-3-030-80285-1_54

119. Watson D, O'Hara MW, Naragon-Gainey K, Koffel E, Chmielewski M, Kotov R, et al. Development and Validation of New Anxiety and Bipolar Symptom Scales for an Expanded Version of the IDAS (the IDAS-II). Assessment. 2012 Dec;19(4):399–420.

120. Wester R, Rubel J, Zimmermann J, Hall M, Kaven L, Watson D. IDAS-II. Inventory of Depression and Anxiety Symptoms - Deutsche Version - Verfahrensdokumentation und Fragebogen. 2021;In Leibniz-Institut für Psychologie (ZPID) (Hrsg.).

121. Bankstahl U, Görtelmeyer R. Measuring subjective complaints of attention and performance failures - development and psychometric validation in tinnitus of the self-assessment scale APSA. Health Qual Life Outcomes. 2013;11(1):86.

122. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction. 2008 Jul 29;24(6):574–94.

123. Brooke J. SUS: A quick and dirty usability scale. Usability Eval Ind. 1995 Nov 30;189.

124. Jacko JA, editor. The human-computer interaction handbook: fundamentals, evolving technologies, and emerging applications. 3rd ed. Boca Raton, FL: CRC Press; 2012. 1452 p. (Human factors and ergonomics).

125. Lewis JR. The System Usability Scale: Past, Present, and Future. International Journal of Human–Computer Interaction. 2018 Jul 3;34(7):577–90.

126. Gao M, Kortum P, Oswald FL. Multi-Language Toolkit for the System Usability Scale. International Journal of Human–Computer Interaction. 2020 Dec 13;36(20):1883–901.

127. Sauro J. A practical guide to the system usability scale: background, benchmarks & best practices. Denver, CO: Measuring Usability LLC; 2011. 162 p.

128. Kothgasser O, Felnhofer A, Hauk N, Kastenhofer E, Gomm J, Kryspin-Exner I. TUI - Technology Usage Inventory Manual. [Internet]. 2012. Available from: https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf

129. Rheinberg F, Vollmeyer R, Burns BD. FAM - Fragebogen zur aktuellen Motivation [Internet]. ZPID (Leibniz Institute for Psychology) – Open Test Archive; 2021 [cited 2025 Dec 26]. Available from: https://www.psycharchives.org/handle/20.500.12034/2281.2

130. Cohen J. Statistical Power Analysis for the Behavioral Sciences [Internet]. 0 ed. Routledge; 2013 [cited 2025 May 14]. Available from: https://www.taylorfrancis.com/books/9781134742707

131. LimeSurvey GmbH. LimeSurvey: An Open Source Survey Tool. [Internet]. 2025. Available from: https://www.limesurvey.org

132. IBM Corp. IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp; 2023.

133. Häussl A, Ehmann E, Pacher A, Knödl K, Huber T, Neundlinger L, et al. Psychological, physical, and social effects of the COVID-19 pandemic on hospital nurses. International Nursing Review. 2021 Dec;68(4):482–92.

134. Kim H, Sefcik JS, Bradway C. Characteristics of Qualitative Descriptive Studies: A Systematic Review. Research in Nursing & Health. 2017 Feb;40(1):23–42.

135. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. 13., überarbeitete Auflage. Weinheim Basel: Beltz; 2022. 148 p.

136. Hussy W, Schreier M, Echterhoff G. Forschungsmethoden in Psychologie und Sozialwissenschaften: für Bachelor. 2. Aufl. Berlin: Springer; 2013. (Springer-Lehrbuch).

137. VERBI Software. MAXQDA 2024 [Internet]. Berlin, Germany: VERBI Software; 2023. Available from: maxqda.com.

138. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on Symptoms of Agitation and Depression in Persons With Dementia Participating in Robot-Assisted Activity: A Cluster-Randomized Controlled Trial. Journal of the American Medical Directors Association. 2015 Oct;16(10):867–73.

139. Beck AT, Steer RA, Brown G. Beck Depression Inventory–II [Internet]. 2011 [cited 2025 Nov 26]. Available from: https://doi.apa.org/doi/10.1037/t00742-000

140. Uher R, Perlis RH, Placentino A, Dernovšek MZ, Henigsberg N, Mors O, et al. SELF-REPORT AND CLINICIAN-RATED MEASURES OF DEPRESSION SEVERITY: CAN ONE REPLACE THE OTHER?: Research Article: Interchangeablility of Self-Report and Clinician-Rated Measures of Depression Severity. Depress Anxiety. 2012 Dec;29(12):1043–9.

141. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med. 2015 Dec;13(1):72.

142. Mavroeides G, Koutra K. Illness representations in depression and their association with clinical and treatment outcomes: A systematic review of the literature. Journal of Affective Disorders Reports. 2021 Apr;4:100099.

143. Zhu X, Chen X, Wu Y, Feng L, Chen X. Longitudinal trajectories of subjective cognitive complaints in patients with major depressive disorder and similar objective cognitive trajectories. BMC Psychiatry. 2025 Mar 26;25(1):287.

144. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis. AJP. 2017 Apr 1;174(4):329–40.

145. Conson M, Zappullo I, Cecere R, Thiebaut De Schotten M, Lauro A, Senese VP, et al. Altercentrism and a change in perspective on the self: The relationships of visuospatial perspective-taking with rumination and mindfulness. Mandolesi L, editor. PLoS One. 2025 Aug 20;20(8):e0316060.

146. Board on Global Health, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. Aging, Functioning, and Rehabilitation: Proceedings of a Workshop [Internet]. Cooper R, Formentos A, Boman A, editors. Washington, D.C.: National Academies Press; 2024 [cited 2025 Jan 2]. Available from: https://nap.nationalacademies.org/catalog/27763

147. Kazantzis N, Whittington C, Dattilio F. Meta-analysis of homework effects in cognitive and behavioral therapy: A replication and extension. Clinical Psychology: Science and Practice. 2010 Jun;17(2):144–56.

148. Cohen ZD, DeRubeis RJ. Treatment Selection in Depression. Annu Rev Clin Psychol. 2018 May 7;14(1):209–36.

149. Høstmælingen A, Ulvenes P, Nissen-Lie HA, Eielsen M, Wampold BE. Comparing outcomes in chronic depression following inpatient psychotherapy for patients continuing

versus discontinuing antidepressant medication. Clin Psychology and Psychoth. 2021 Sep;28(5):1111–27.

150. Robinson NL, Connolly J, Suddrey G, Kavanagh DJ. A Brief Wellbeing Training Session Delivered by a Humanoid Social Robot: A Pilot Randomized Controlled Trial. Int J of Soc Robotics. 2024 May;16(5):937–51.

151. Fasola J, Matarić MJ. Socially Assistive Robot Exercise Coach: Motivating Older Adults to Engage in Physical Exercise. In: Desai JP, Dudek G, Khatib O, Kumar V, editors. Experimental Robotics [Internet]. Heidelberg: Springer International Publishing; 2013 [cited 2025 Nov 11]. p. 463–79. (Springer Tracts in Advanced Robotics; vol. 88). Available from: https://link.springer.com/10.1007/978-3-319-00065-7_32

152. Juras L, Sedlar MK, Vrhovski G, Vranic A. N-Back training in middle adulthood: evidence for transfer only to structurally similar task. Curr Psychol. 2024 Jun;43(21):19110–20.

153. Chatzoglou PD, Lazaraki V, Apostolidis SD, Gasteratos AC. Factors Affecting Acceptance of Social Robots Among Prospective Users. Int J of Soc Robotics. 2024 Jun;16(6):1361–80.

154. Mandl S, Brade J, Bretschneider M, Asbrock F, Meyer B, Jahn G, et al. Perception of embodied digital technologies: robots and telepresence systems. Hum-Intell Syst Integr. 2023 Nov;5(1–2):43–62.

155. Chiauzzi E, Williams A, Mariano TY, Pajarito S, Robinson A, Kirvin-Quamme A, et al. Demographic and clinical characteristics associated with anxiety and depressive symptom outcomes in users of a digital mental health intervention incorporating a relational agent. BMC Psychiatry. 2024 Jan 30;24(1):79.

156. Belsky J. Differential Susceptibility to Environmental Influences. ICEP. 2013 Nov;7(2):15–31.

157. Kuehner C. Why is depression more common among women than among men? The Lancet Psychiatry. 2017 Feb;4(2):146–58.

158. Trivedi MH, Greer TL. Cognitive dysfunction in unipolar depression: Implications for treatment. Journal of Affective Disorders. 2014 Jan;152–154:19–27.

159. Bakbani-Elkayam S, Dolev-Amit T, Hemed E, Zilcha-Mano S, Eitam B. Intact modulation of response vigor in major depressive disorder. Motiv Emot. 2024 Apr;48(2):209–21.

160. Saberi S, Ahmadi R, Khakpoor S, Pirzeh R, Hasani M, Moradveisi L, et al. Comparing the effectiveness of behavioral activation in group vs. self-help format for reducing depression, repetitive thoughts, and enhancing performance of patients with major depressive disorder: a randomized clinical trial. BMC Psychiatry. 2024 Jul 19;24(1):516.

161. Snyder HR. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. Psychological Bulletin. 2013;139(1):81–132.

162. Lampit A, Hallock H, Valenzuela M. Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. Gandy S, editor. PLoS Med. 2014 Nov 18;11(11):e1001756.

163. Koster EHW, Hoorelbeke K, Onraedt T, Owens M, Derakshan N. Cognitive control interventions for depression: A systematic review of findings from training studies. Clinical Psychology Review. 2017 Apr;53:79–92.

164. Hirnstein M, Stuebs J, Moè A, Hausmann M. Sex/Gender Differences in Verbal Fluency and Verbal-Episodic Memory: A Meta-Analysis. Perspect Psychol Sci. 2023 Jan;18(1):67–90.

165. Hofer G, Hünninghaus M, Platzer J, Grinschgl S, Neubauer A. Women's Humility and Men's Lack of Hubris: Gender Biases in Self-Estimated Spatial Intelligence. Sex Roles. 2025 Apr;91(4):29.

166. DeCasien AR, Guma E, Liu S, Raznahan A. Sex differences in the human brain: a roadmap for more careful analysis and interpretation of a biological reality. Biol Sex Differ. 2022 Jul 26;13(1):43.

167. Carvalho Silva R, Pisanu C, Maffioletti E, Menesello V, Bortolomasi M, Gennarelli M, et al. Biological markers of sex-based differences in major depressive disorder and in antidepressant response. European Neuropsychopharmacology. 2023 Nov;76:89–107.

168. Yagi S, Galea LAM. Sex differences in hippocampal cognition and neurogenesis. Neuropsychopharmacology. 2019 Jan;44(1):200–13.

169. Arenaza-Urquijo EM, Boyle R, Casaletto K, Anstey KJ, Vila-Castelar C, Colverson A, et al. Sex and gender differences in cognitive resilience to aging and Alzheimer's disease. Alzheimer's & Dementia. 2024 Aug;20(8):5695–719.

170. Bosch R, Noordzij ML, Westerhof GJ, Bohlmeijer ET. Integration of Sensor-Based Technology in Mental Healthcare: A Systematic Scoping Review. J technol behav sci. 2024 Jul 2;10(1):201–17.

171. Aravind A, Agarwal M, Malhotra S, Ayyub S. Effectiveness of Acceptance and Commitment Therapy on Mental Health Issues: A Systematic Review. Annals of Neurosciences. 2025 Oct;32(4):321–7.

172. Braun P, Drüge M, Hennemann S, Nitsch FJ, Staeck R, Apolinário-Hagen J. Acceptance of E-Mental Health Services for Different Application Purposes Among Psychotherapists in Clinical Training in Germany and Switzerland: Secondary Analysis of a Cross-Sectional Survey. Front Digit Health. 2022 Feb 28;4:840869.

173. Van Deursen AJ, Van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media & Society. 2019 Feb;21(2):354–75.

174. Boileau L, Gaudreau P, Pétrin-Pomerleau P. Fear of Failure as Motivation: A Novel Conceptualization and Measure. Journal of Personality Assessment. 2025 Sep 19;1–15.

175. Chung JOK, Li WHC, Ho KY, Lam KKW, Cheung AT, Ho LLK, et al. Adventure-based training to enhance resilience and reduce depressive symptoms among juveniles: A randomized controlled trial. Res Nurs Health. 2021 Jun;44(3):438–48.

176. Duru E, Balkis M, Duru S. Fear of failure and academic satisfaction: the mediating role of emotion regulation difficulties and procrastination. Eur J Psychol Educ. 2024 Sep;39(3):2901–14.

177. King RB. Is a performance- avoidance achievement goal always maladaptive? Not necessarily for collectivists. Personality and Individual Differences. 2016 Sep;99:190–5.

178. Wimmer S, Lackner HK, Papousek I, Paechter M. Goal Orientations and Activation of Approach Versus Avoidance Motivation While Awaiting an Achievement Situation in the Laboratory. Front Psychol. 2018 Aug 28;9:1552.

179. Chen Z. Cognitive Biases and Test Anxiety: The Impact of Negative Thinking Patterns on Academic Performance. LNEP. 2024 Nov 29;67(1):96–100.

180. Murman DL. The Impact of Age on Cognition. Semin Hear. 2015 Aug;36(3):111–21.

181. Buechner VL, Pekrun R, Lichtenfeld S. The Achievement Pride Scales (APS). European Journal of Psychological Assessment. 2018 May;34(3):181–92.

182. Angulo J, Fleury V, Péron JA, Penzenstadler L, Zullino D, Krack P. Shame in Parkinson'S Disease: A Review. J Parkinsons Dis. 2019;9(3):489–99.

183. Bynum WE, Artino AR, Uijtdehaage S, Webb AMB, Varpio L. Sentinel Emotional Events: The Nature, Triggers, and Effects of Shame Experiences in Medical Residents. Academic Medicine. 2019 Jan;94(1):85–93.

184. Oren-Yagoda R, Rosenblum M, Aderka IM. Gender Differences in Shame Among Individuals with Social Anxiety Disorder. Cogn Ther Res. 2024 Aug;48(4):720–9.

185. Streck H, Nishen AK, Kessels U. Instrumentality Gives Girls the Edge: Gender-Differential Relations Between Instrumentality, Achievement Motivation, and Self-Esteem. Sex Roles. 2022 Mar;86(5–6):379–94.

186. Bandura A. Self efficacy: the exercise of control. New York (N.Y.): W. H. Freeman; 1997.

187. Kühne K, Fischer MH, Zhou Y. The Human Takes It All: Humanlike Synthesized Voices Are Perceived as Less Eerie and More Likable. Evidence From a Subjective Ratings Study. Front Neurorobot. 2020 Dec 16;14:593732.

188. Deci EL, Ryan RM. The Empirical Exploration of Intrinsic Motivational Processes. In: Advances in Experimental Social Psychology [Internet]. Elsevier; 1980 [cited 2025 Nov 26]. p. 39–80. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0065260108601306

189. Morris LS, Grehl MM, Rutter SB, Mehta M, Westwater ML. On what motivates us: a detailed review of intrinsic *v.* extrinsic motivation. Psychol Med. 2022 Jul;52(10):1801–16.

190. Allotey J, Marlin N. Well-conducted randomisation will improve balance and reliability of treatment estimates: Randomised controlled trials: random acts of allocation. BJOG. 2019 Sep;126(10):1287–1287.

191. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? BMJ. 1999 May 1;318(7192):1209.

192. Sedgwick P. Why randomise in clinical trials? BMJ. 2012 Aug 22;345(aug22 1):e5584–e5584.

193. Šabanović S, Chang WL, Bennett CC, Piatt JA, Hakken D. A Robot of My Own: Participatory Design of Socially Assistive Robots for Independently Living Older Adults Diagnosed with Depression. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population Design for Aging [Internet]. Cham: Springer International Publishing; 2015 [cited 2025 Mar 3]. p. 104–14. (Lecture Notes in Computer Science; vol. 9193). Available from: http://link.springer.com/10.1007/978-3-319-20892-3_11

194. Eyssel F, Hegel F. (S)he's Got the Look: Gender Stereotyping of Robots[1]. J Applied Social Pyschol. 2012 Sep;42(9):2213–30.

195. Huijnen CA, As Lexis M, De Witte LP. Robots as New Tools in Therapy and Education for Children with Autism. Int J Neurorehabilitation Eng [Internet]. 2017 [cited 2026 Jan 8];04(04). Available from: https://www.omicsonline.org/open-access/robots-as-new-tools-in-therapy-and-education-for-children-with-autism-2376-0281-1000278.php?aid=92196

196. Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. Clinical Psychology Review. 2015 Feb;35:35–46.

197. Lopez-Alcalde J, Wieland LS, Yan Y, Barth J, Khami MR, Shivalli S, et al. Methodological Challenges in Randomized Controlled Trials of mHealth Interventions: Cross-Sectional Survey Study and Consensus-Based Recommendations. J Med Internet Res. 2024 Dec 19;26:e53187.

198. Krukowski RA, Ross KM, Western MJ, Cooper R, Busse H, Forbes C, et al. Digital health interventions for all? Examining inclusivity across all stages of the digital health intervention research process. Trials. 2024 Jan 30;25(1):98.

199. Tapus A, Tapus C, Mataric MJ. The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia. In: 2009 IEEE International Conference on Rehabilitation Robotics [Internet]. Kyoto, Japan: IEEE; 2009 [cited 2024 Dec 27]. p. 924–9. Available from: http://ieeexplore.ieee.org/document/5209501/

200. Sharkey A, Sharkey N. We need to talk about deception in social robotics! Ethics Inf Technol. 2021 Sep;23(3):309–16.

7. Appendix

7.1 Tables and Figures

7.1.1 IDAS-II – Subscale Depression

Table 8: *Depression Scores of the Inventory of Depression and Anxiety Symptoms - German Version (IDAS-II) at Baseline (t_0) and Post-Intervention (t_1) by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

IDAS-II Sub. Dep	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI*	
							Lower	Upper
SAR[a] at t_0	16	56.31	12.47	.014	30	.176	-2.969	15.469
Tablet[b] at t_0	16	50.06	13.05					
SAR[a] at t_1	16	48.75	13.59	.663	30	.488	-12.213	5.963
Tablet[b] at t_1	16	51.88	11.48					
Female at t_0	16	56.75	13.15	.101	30	.121	-2.005	16.255
Male at t_0	16	49.63	12.11					
Female at t_1	16	53.69	13.56	.448	30	.128	-2.060	15.560
Male at t_1	16	46.94	10.66					
≤ 30 years at t_0	16	51.19	14.41	2.414	30	.391	-13.391	5.391
≥ 31 years at t_0	16	55.19	11.42					
≤ 30 years at t_1	16	47.25	12.69	.660	30	.169	-14.998	2.748
≥ 31 years at t_1	16	53.38	11.871					

Note: t-test. [a] = with motivational feedback. [b] = without motivational feedback.
Dep. = Depression. **df** = Degrees of freedom. **F** = F-statistics. **IDAS-II** = Inventory of Depression and Anxiety Symptoms – German Version. **M** = Mean. **n** = Sample size of a specific group within the total sample size. **SAR** = Socially Assistive Robot. **SD** = Standard deviation. **Sub.** = Subscale. **t_0** = Time of Measurement 1. **t_1** = Time of Measurement 2. **vs**. = versus.
* p < .05. ** p < .01. *** p < .001.

7.1.2 Attention and Performance Self-Assessment (APSA).

Table 9: *Overall Score and Scores of the Subscales for Attention and Performance Self-Assessment (APSA) at Baseline (t_0) and Post-Intervention (t_1) by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

APS-20	**n**	***M***	***SD***	***F***	***df.***	***Two-sided p***	***95% CI***	
							Lower	**Upper**
SAR[a] at t_0	16	61.19	11.90	1.516	30	.407	-5.973	14.348
Tablet[b] at t_0	16	57.00	15.94					
SAR[a] at t_1	16	56.00	12.73	.542	30	.240	-15.217	3.967
Tablet[b] at t_1	16	61.63	13.80					
Female at t_0	16	67.56	10.27	.135	30	**<.001*** **	8.826	25.049
Male at t_0	16	50.63	12.11					
Female at t_1	16	63.56	15.14	1.061	30	.043*	.343	18.657
Male at t_1	16	54.06	9.60					
≤ 30 years at t_0	16	61.63	11.55	1.134	30	.314	-5.043	15.168
≥ 31 years at t_0	16	56.56	16.07					
≤ 30 years at t_1	16	57.94	14.38	.055	30	.718	-11.547	8.047
≥ 31 years at t_1	16	59.69	12.72					

AP-F1	**n**	***M***	***SD***	***F***	***df.***	***Two-sided p***	***95% CI***	
							Lower	**Upper**
SAR[a] at t_0	16	25.69	6.43	1.350	30	.645	-4.028	6.403
Tablet[b] at t_0	16	24.50	7.93					
SAR[a] at t_1	16	23.81	5.98	.487	30	.225	-7.287	1.787
Tablet[b] at t_1	16	26.56	6.57					
Female at t_0	16	29.25	5.96	.001	30	**<.001*** **	4.094	12.531
Male at t_0	16	20.94	5.721					
Female at t_1	16	27.00	6.92	.762	30	.107	-.826	8.076
Male at t_1	16	23.38	5.29					
≤ 30 years at t_0	16	26.69	6.58	.297	30	.211	-1.910	8.285
≥ 31 years at t_0	16	23.50	7.51					
≤ 30 years at t_1	16	25.69	6.27	.159	30	.663	-3.636	5.636
≥ 31 years at t_1	16	24.69	6.56					

AP-F2	**n**	***M***	***SD***	***F***	***df.***	***Two-sided p***	***95% CI***	
							Lower	**Upper**
SAR[a] at t_0	16	29.09	5.75	.604	30	.241	-1.945	7.445
Tablet[b] at t_0	16	26.31	7.17					
SAR[a] at t_1	16	26.31	6.72	.000	30	.379	-6.781	2.656
Tablet[b] at t_1	16	28.38	6.34					
Female at t_0	16	31.31	5.12	.026	30	**<.001*** **	3.277	11.223
Male at t_0	16	24.06	5.85					
Female at t_1	16	29.94	7.12	.118	30	**.022***	.815	9.560

Male at t_1	16	24.75	4.75					
≤ 30 years at t_0	16	28.25	5.09	1.307	30	.636	-3.662	5.912
≥ 31 years at t_0	16	27.13	7.87					
≤ 30 years at t_1	16	26.06	7.10	.478	30	.273	-7.247	2.122
≥ 31 years at t_1	16	28.63	5.87					

Note: t-test. [a] = with motivational feedback. [b] = without motivational feedback.
AP-F1 = Difficulties Prospective Memory Performance. **AP-F2** = Difficulties in Maintaining Focused Attention. **APS-20** = Overall Score of the Attention and Performance Self-Assessment. **APSA** = Attention and Performance Self-Assessment. **df** = Degrees of freedom. **F** = F-statistics. **M** = Mean. **n** = Sample size of a specific group within the total sample size. **SAR** = Socially Assistive Robot. **SD** = Standard deviation. **ToM** = Time of Measurement. **t_0** = Time of Measurement 1. **t_1** = Time of Measurement 2. **vs**. = versus.
***** p < .05. ****** p < .01. ******* p < .001.

Note: AP-F1 = Difficulties Prospective Memory Performance. **AP-F2** = Difficulties in Maintaining Focused Attention. **APS-20** = Overall Score of the Attention and Performance Self-Assessment. **APSA** = Attention and Performance Self-Assessment. **SAR** = Socially Assistive Robot. **t_0** = Time of Measurement 1. **t_1** = Time of Measurement 2. **vs**. = versus.

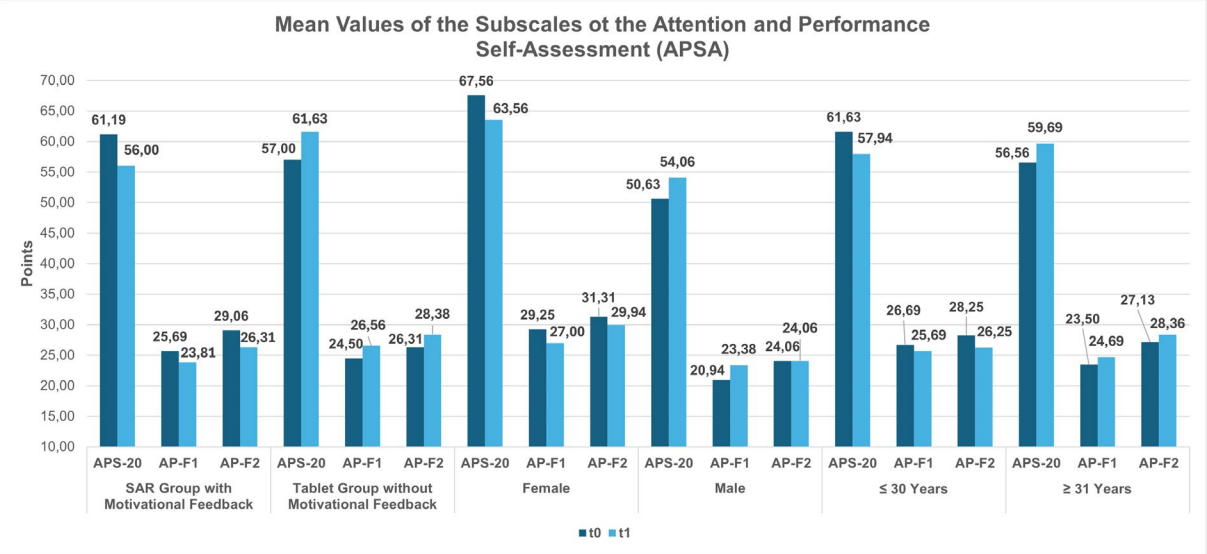

Figure 6: *Mean Values of the Subscales of the Attention and Performance Self-Assessment (APSA) at Baseline (t_0) and Post-Intervention (t_1) by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years)*

7.1.3 Technology Acceptance Model (TAM) Questionnaire

Table 10: *Subscale Scores of the Technology Acceptance Model (TAM) Questionnaire by Age (≤ 30 Years vs. ≥ 31 Years).*

Variable	≤ 30 Years (*n* = 16)		≥ 31 Years (*n* = 16)		Mean Difference	*F*	Sig.[a]	η^2_p	95% CI	
									Lower	Upper
TAM PU	31.50	4.71	31.69	3.30	-.691	.205	.654	.007	-3.818	2.436
TAM PEU	33.56	4.45	31.50	4.14	2.808	3.180	.085	.102	-.417	6.034
TAM ITU	11.61	1.50	11.25	1.77	.512	.703	.409	.025	-.738	1.762

Note: MANCOVA. [a] = adjustment for multiple comparisons: Bonferroni

CI = confidence Interval. **F =** F-test. **ITU** = Intention to Use. **M** = Mean. **n** = Sample size of a specific group within the total sample size. η^2_p = Partial Eta Squared. **PEU** = Perceived Ease of Use. **PU** = Perceived Usefulness. **SD** = Standard deviation. **Sig.** = Significance. **TAM** = Technology Acceptance Model.
* p < .05. ** p < .01. *** p < .001.

7.1.4 System Usability Scale (SUS)

Table 11: *System Usability Scale (SUS) Scores by Age (≤ 30 Years vs. ≥ 31 Years).*

Variable	≤ 30 Years (*n* = 16)		≥ 31 Years (*n* = 16)		Mean Difference	*F*	Sig.[a]	η^2_p	95% CI	
	M	*SD*	*M*	*SD*					Lower	Upper
SUS	70.78	14.45	68.44	11.06	3.955	.629	.435	.022	-5.863	13.268

Note: ANCOVA. [a] = adjustment for multiple comparisons: Bonferroni

CI = confidence Interval. **F =** F-test. **M** = Mean. **n** = Sample size of a specific group within the total sample size. η^2_p = Partial Eta Squared. **SD** = Standard deviation. **Sig.** = Significance. **SUS** = System Usability Scale. **vs**. = versus.
* p < .05. ** p < .01. *** p < .001.

7.1.5 Technology Usage Inventory (TUI)

Table 12: *Subscale Scores of the Technology Usage Inventory (TUI) by Age (≤ 30 Years vs. ≥ 31 Years).*

Variable	≤ 30 Years (*n* = 16)		≥ 31 Years (*n* = 16)		Mean Difference	*F*	Sig.[a]	η^2_p	95% CI	
	M	*SD*	*M*	*SD*					Lower	Upper
TUI OVE [1]	106.06	13.76	95.06	14.18	13.127	6.306	**0,18***	.184	99.742	114.511
TUI INT [2]	20.56	4.24	14.75	4.62	6.664	16.747	**< .001*** **	.374	3.328	10.000
TUI USE [2]	17.00	6.06	16.86	6.63	-.202	.007	.934	.000	-5.195	4.790
TUI SKE [2]	11.56	4.83	12.69	5.19	-.886	.226	.638	.008	-4.703	2.931
TUI ANX [2]	10.94	4.75	12.19	6.11	-1.302	.457	.504	.016	-5.247	2.642
TUI UF [2]	17.12	3.09	15.06	3.95	2.954	5.517	**.026***	.165	.378	5.530
TUI ACC [2]	8.56	3.46	9.31	4.15	-1.387	.960	.335	.033	-4.286	1.512
TUI CUR [2]	19.88	5.08	14.62	5.17	5.922	9.402	**.005****	251	1.966	9.879
TUI ITU [2]	205.19	51.38	186.19	69.24	31.86	2.214	.148	.073	-75.726	11.999

Note: [1] = ANCOVA; [2] = MANCOVA. [a] = adjustment for multiple comparisons: Bonferroni
ACC= Accessibility. **ANX** = Anxiety. **CI** = confidence Interval. **CUR** = Curiosity. **F =** F-test. **INT** = Interest. **ITU** = Intention to Use. **M** = Mean. **n** = Sample size of a specific group within the total sample size. η^2_p = Partial Eta Squared. **SD** = Standard deviation. **Sig.** = Significance. **SKE** = Skepticism. **TUI** = Technology Usage Inventory. **OVE** = overall. **UF** = User-Friendliness. **USE** = Usefulness.
* p < .05. ** p < .01. *** p < .001.

7.1.6 Questionnaire on Current Motivation (QCM)

Table 13: *Subscale Scores of the Questionnaire on Current Motivation (QCM) at Baseline (t_0) and Post-Intervention (t_1) by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

QCM – FoF	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI*	
							Lower	Upper
SAR[a] at t_0	16	3.40	1.81	6.429	30	.619	-.804	1.329
Tablet[b] at t_0	16	3.13	1.03					
SAR[a] at t_1	16	3.61	1.76	.621	30	.632	-.884	1.434
Tablet[b] at t_1	16	3.33	1.43					
Female at t_0	16	3.65	1.51	.465	30	.142	-.270	1.795
Male at t_0	16	2.88	1.33					
Female at t_1	16	3.80	1.40	1.289	30	.253	-.488	1.788
Male at t_1	16	3.15	1.73					
≤ 30 years at t_0	16	2.65	1.17	1.587	30	**.014***	-2.204	-.270
≥ 31 years at t_0	16	3.88	1.48					
≤ 30 years at t_1	16	3.05	1.61	3.77	30	.132	-1.970	.270
≥ 31 years at t_1	16	3.90	1.48					

QCM – CHAL	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI*	
							Lower	Upper
SAR[a] at t_0	16	4.67	1.60	24.232	30	.526	-.613	1.176
Tablet[b] at t_0	16	4.39	.69					
SAR[a] at t_1	16	4.42	1.18	3.126	30	**.034***	-1.561	-.063
Tablet[b] at t_1	16	5.23	.86					
Female at t_0	16	4.57	1.33	.345	30	.833	-.806	.994
Male at t_0	16	4.48	1.15					
Female at t_1	16	4.95	.94	1.029	30	.529	-.552	1.052
Male at t_1	16	4.70	1.25					
≤ 30 years at t_0	16	4.09	1.21	.192	30	**.042***	-1.714	-.035
≥ 31 years at t_0	16	4.96	1.11					
≤ 30 years at t_1	16	4.57	1.39	10.146	30	.204	-1.286	.286
≥ 31 years at t_1	16	5.07	.65					

QCM - PoS	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI*	
							Lower	Upper
SAR[a] at t_0	16	4.84	1.51	.880	30	.786	-.805	1.055
Tablet[b] at t_0	16	4.71	1.00					
SAR[a] at t_1	16	5.53	1.27	1.976	30	.442	-.506	1.131
Tablet[b] at t_1	16	5.21	.97					
Female at t_0	16	4.50	1.32	.049	30	.216	-1.470	.345
Male at t_0	16	5.06	1.18					
Female at t_1	16	5.31	1.18	.636	30	.759	-950	.700

Male at t_1	16	5.43	1.09					
≤ 30 years at t_0	16	5.06	1.08	.193	30	.216	-.345	1.470
≥ 31 years at t_0	16	4.50	1.08					
≤ 30 years at t_1	16	5.60	1.14	.061	30	.246	-.339	1.277
≥ 31 years at t_1	16	5.14	1.09					

QCM - INT	**n**	***M***	***SD***	***F***	***df.***	***Two-sided p***	***95% CI***	
							Lower	**Upper**
SAR[a] at t_0	16	4.67	1.34	.006	30	.119	-.205	1.705
Tablet[b] at t_0	16	3.92	1.30					
SAR[a] at t_1	16	4.95	1.28	2.706	30	.419	-1.090	.465
Tablet[b] at t_1	16	5.26	.82					
Female at t_0	16	4.43	1.49	.461	30	.566	-708	1.271
Male at t_0	16	4.15	1.23					
Female at t_1	16	5.46	.98	.207	30	.056	-.020	1.458
Male at t_1	16	4.75	1.06					
≤ 30 years at t_0	16	4.34	1.53	1.601	30	.849	-.901	1.088
≥ 31 years at t_0	16	4.25	1.20					
≤ 30 years at t_1	16	5.29	.92	1.181	30	.330	-.399	1.149
≥ 31 years at t_1	16	4.92	1.20					

Note: t-test. [a] = with motivational feedback. [b] = without motivational feedback. * p < .05. ** p < .01. *** p < .001.
CHAL = Challenge. **df** = Degrees of freedom. **F** = F-statistics. **FoF** = Fear of Failure. INT = Interest. **M** = Mean. **N** = Sample size of a specific group within the total sample size. **PoS** = Probability of Success. **QCM** = Questionnaire on Current Motivation. **SAR** = Socially Assistive Robot. **SD** = Standard deviation. **t_0** = Time of Measurement 1. **t_1** = Time of Measurement 2. **vs**. = versus.
* p < .05. ** p < .01. *** p < .001.

7.1.7 Questionnaire on Achievement Motivation (QAM)

Note: QAM = Questionnaire on Achievement Motivation. **SAR** = Socially Assistive Robot. $\mathbf{t_0}$ = Time of Measurement 1. $\mathbf{t_1}$ = Time of Measurement 2. **vs**. = versus.

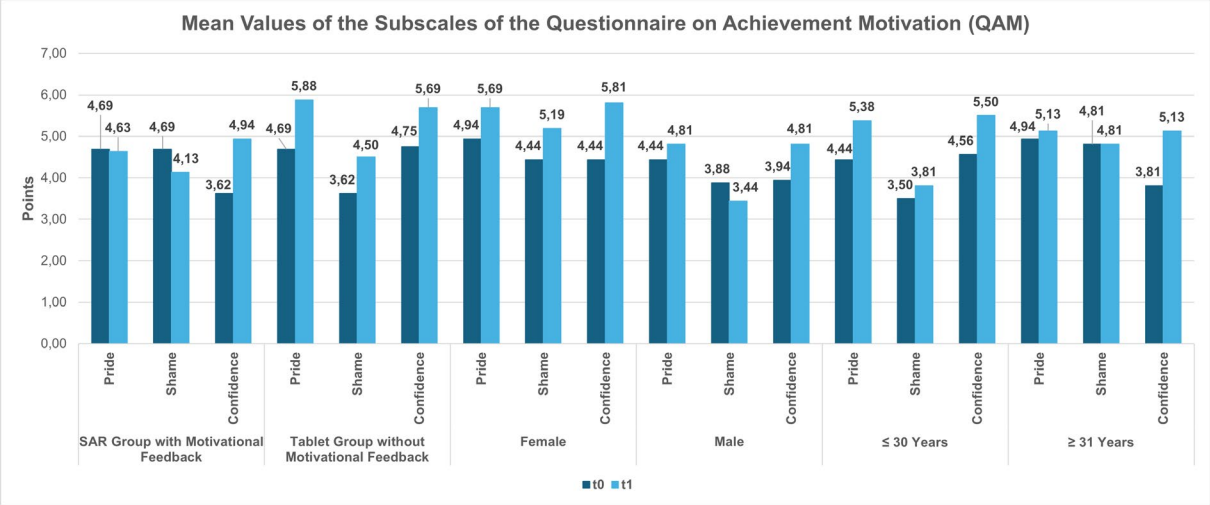

Figure 7: *Mean Values of the Subscales of the Questionnaire on Achievement Motivation (QAM) at Baseline (t_0) and Post-Intervention (t_1) by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years)*

Table 14: *Subscale Scores of the Questionnaire on Achievement Motivation (QAM) at Baseline (t_0) and Post-Intervention (t_1) by Group (SAR Group with Motivational Feedback vs. Tablet Group without Motivational Feedback), Sex (Female vs. Male) and Age (≤ 30 Years vs. ≥ 31 Years).*

QAM – PRI	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI* Lower	*95% CI* Upper
SAR[a] at t_0	16	4.69	1.92	0.000	30	1.00	-1.376	1.376
Tablet[b] at t_0	16	4.69	1.88					
SAR[a] at t_1	16	4.63	1.74	3.803	30	**.019***	-2.284	-.216
Tablet[b] at t_1	16	5.88	1.02					
Female at t_0	16	4.94	1.98	.429	30	.460	-.863	1.863
Male at t_0	16	4.44	1.78					
Female at t_1	16	5.69	1.70	.980	30	.110	-.211	1.961
Male at t_1	16	4.81	1.27					
≤ 30 years at t_0	16	4.44	2.06	.645	30	.460	-1.863	.863
≥ 31 years at t_0	16	4.94	1.69					
≤ 30 years at t_1	16	5.38	1.82	1.358	30	.655	-.880	1.380
≥ 31 years at t_1	16	5.13	1.25					

QAM – SHA	n	*M*	*SD*	*F*	*df.*	*Two-sided p*	*95% CI* Lower	*95% CI* Upper
SAR[a] at t_0	16	4.69	2.38	.004	30	.221	-.672	2.797
Tablet[b] at t_0	16	3.63	2.41					
SAR[a] at t_1	16	4.13	2.30	.002	30	.645	-2.020	1.270
Tablet[b] at t_1	16	4.50	2.25					

Female at t_0	16	4.44	2.44	.000	30	.521	-1.205	2.330
Male at t_0	16	3.88	2.44					
Female at t_1	16	5.19	2.13	.037	30	**.025***	.233	3.267
Male at t_1	16	3.44	2.06					
≤ 30 years at t_0	16	3.50	2.39	.039	30	.128	-3.023	.398
≥ 31 years at t_0	16	4.81	2.34					
≤ 30 years at t_1	16	3.81	2.37	.791	30	.214	-2.608	-.608
≥ 31 years at t_1	16	4.81	2.07					

QAM – CON	**n**	***M***	***SD***	***F***	***df.***	***Two-sided p***	***95% CI***	
							Lower	**Upper**
SAR[a] at t_0	16	3.63	1.85	.148	30	.077	-2.382	.132
Tablet[b] at t_0	16	4.75	1.61					
SAR[a] at t_1	16	4.94	1.87	1.838	30	.216	-1.961	.461
Tablet[b] at t_1	16	5.69	1.44					
Female at t_0	16	4.44	1.86	.787	30	.442	-.811	1.811
Male at t_0	16	3.94	1.76					
Female at t_1	16	5.81	1.75	.204	30	.095	.580	2.186
Male at t_1	16	4.81	1.51					
≤ 30 years at t_0	16	4.56	1.99	1.299	30	.246	-.544	2.044
≥ 31 years at t_0	16	3.81	1.55					
≤ 30 years at t_1	16	5.50	1.82	.505	30	.540	-.860	1.610
≥ 31 years at t_1	16	5.13	1.58					

Note: t-test. [a] = with motivational feedback. [b] = without motivational feedback. * p < .05. ** p < .01. *** p < .001.
CON = Confidence. **df** = Degrees of freedom. **F** = F-statistics. **M** = Mean. **N** = Sample size of a specific group within the total sample size. **PRI** = Pride. **QAM** = Questionnaire on Achievement Motivation. **SAR** = Socially Assistive Robot. **SD** = Standard deviation. **SHA** = Shame. **Sig.** = Significance level for ANOVA. t_0 = Time of Measurement 1. t_1 = Time of Measurement 2. **vs**. = versus.
* p < .05. ** p < .01. *** p < .001.

7.2 Non standardized questionnaire

Table 15: *Sociodemographic questionnaire*

Participant Code	……..
Gender	☐ Female ☐ Male ☐ Divers ☐ Inter ☐ Open ☐ „No Response“ ☐ Other: ……
In which year were you born?	……
In which country were you born?	…… (Drop-down-list of all countries)
What nationality or nationalities do you hold?	1. …… 2. ……
Did you grow up in Austria?	☐ Yes ☐ No If no, where are you from? …… If no, how long have you lived in Austria? ☐ less than 5 Jahre ☐ 5 – 10 years ☐ 11 – 15 years ☐ more than 15 years
Do you live in Styria?	☐ Yes ☐ No
If yes, in which district do you live?	☐ Bruck-Mürzzuschlag ☐ Liezen ☐ Deutschlandsberg ☐ Murau ☐ Graz ☐ Murtal ☐ Graz-Umgebung ☐ South-East-Styria ☐ Hartberg-Fürstenfeld ☐ Voitsberg ☐ Leibnitz ☐ Weiz ☐ Leoben ☐ Other district: ……
If no, in which federal state do you live?	☐ Vienna ☐ Carithia ☐ Burgenland ☐ Salzburg ☐ Upper Austria ☐ Tyrol ☐ Lower Austria ☐ Vorarlberg
If no, in which district do you live?	……
Do you live in the district capital?	☐ Yes ☐ No
Do you live in a city or in the countryside?	☐ City ☐ Countryside
What is your highest level of completed education?	☐ Primary School ☐ Compulsory School ☐ Apprenticeship without master craftsman’s examination ☐ Apprenticeship with master craftsman’s examination ☐ Polytechnic School ☐ Vocational School (Apprenticeship + A-levels) ☐ General / Middle School without A-levels ☐ General / Middle School with A-levels ☐ Academy, College ☐ Courses (e.g., Universities, Universities of Applied Sciences, Teacher Training Colleges) ☐ University or University of Applied Sciences ☐ Teacher Training College ☐ Other: _____
What is your current or previous occupation?	☐ Retired ☐ Employee ☐ Homemaker ☐ Farmer ☐ Civil Servant ☐ Self-employed ☐ Worker ☐ Not employed
Marital status	☐ Single ☐ Married, living together ☐ In a relastionship, living together ☐ Married, seperated

	☐ In a relastionship, not living together ☐ Divorced ☐ Widowed
Number of children	☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ >10
Age of your first child?	……
Age of your second child?	……
...	….
Age of your tenth child?	……

Questionnaire on Achievement Motivation (QAM)

Table 16: *Questionnaire on Achievement Motivation*

		high						low
		1	**2**	**3**	**4**	**5**	**6**	**7**
1	How great would your pride be if you succeeded in cognitive training?							
2	How great would your shame be if you did not succeed in cognitive training?							
3	How confident are you that you will succeed in cognitive training in the future?							

Participant Feedback on Cognitive Training Experience

Table 17: Participant Feedback on Cognitive Training Experience

Questions	Answers
What aspects of the cognitive training did you find most helpful or enjoyable?	
Were there any parts of the training you found difficult or unhelpful? Please describe.	
How did you experience interacting with the SAR “Pepper” during the training?	
What would you suggest to improve the training sessions in the future?	
Did the training meet your personal expectations? Why or why not?	
How did the training influence your motivation or mood?	
Were there any technical or practical issues that affected your experience?	
What did you like most about the overall setup of the study?	
What would you change to make the training more engaging or effective?	
Do you have any other comments, suggestions, or feedback you would like to share?	

7.3 Publication

Parts of this thesis have been published in the following article by Haeussl et al. (112) with permission of publisher (Frontiers in Psychiatry – Digital Mental Health).

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 15 January 2026
DOI 10.3389/fpsyt.2025.1661730

OPEN ACCESS

EDITED BY
Guang Chen,
The University of Hong Kong, Hong Kong SAR, China

REVIEWED BY
Lizhou Fan,
Harvard Medical School, United States
Jiayan Zhou,
Stanford University, United States

*CORRESPONDENCE
Nina Dalkner
nina.dalkner@medunigraz.at

RECEIVED 08 July 2025
REVISED 13 November 2025
ACCEPTED 22 December 2025
PUBLISHED 15 January 2026

CITATION
Haeussl A, Zwigl I, Stojec L, Smolak I, Stijic M, Stross T, Lenger M, Fellendorf FT, Guggemos S, Schoenthaler EMD, Pszeida M, Orgel T, Draxler S, Schneeberger M, Russegger S, Zuschnegg J, Steindl D, Schultz A, Schuessler S, Macher M, Paletta L, Dalkner N and Reininghaus EZ (2026) Evaluating usability and acceptance of a socially assistive robot supported cognitive training for depression – results of the randomized controlled pilot study 'AMIGA'. *Front. Psychiatry* 16:1661730. doi: 10.3389/fpsyt.2025.1661730

Evaluating usability and acceptance of a socially assistive robot supported cognitive training for depression – results of the randomized controlled pilot study 'AMIGA'

Alfred Haeussl[1,2], Ina Zwigl[1], Lena Stojec[1], Irina Smolak[1], Marko Stijic[3], Tatjana Stross[1], Melanie Lenger[1], Frederike T. Fellendorf[1], Suher Guggemos[1], Elena M. D. Schoenthaler[1], Martin Pszeida[4], Thomas Orgel[4], Sandra Draxler[4], Michael Schneeberger[4], Silvia Russegger[4], Julia Zuschnegg[5], Dominik Steindl[5], Anna Schultz[5], Sandra Schuessler[5], Michael Macher[6,7], Lucas Paletta[4], Nina Dalkner[1]* and Eva Z. Reininghaus[1]

[1]Department of Psychiatry, Psychosomatic and Psychotherapy, Clinical Division of Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Graz, Austria, [2]Department of Healthcare and Nursing, Department of Health Sciences, Faculty of Health and Social Sciences, USTP – University of Applied Sciences St. Poelten, Sankt Poelten, Austria, [3]Department of Neurology, Division of Neurogeriatrics, Medical University of Graz, Graz, Austria, [4]DIGITAL – Institute for Digital Technologies, Joanneum Research Forschungsgesellschaft mbH, Graz, Austria, [5]Institute of Social Medicine and Epidemiology, Medical University of Graz, Graz, Austria, [6]Department of Media and Digital Technologies, Faculty of Engineering and Business, USTP – University of Applied Sciences St. Poelten, Sankt Poelten, Austria, [7]Humanizing Technologies GmbH, Vienna, Austria

The integration of socially assistive robots (SARs) in mental healthcare offers promising opportunities to enhance traditional treatments. The humanoid SAR "Pepper" has shown potential for supporting cognitive and emotional interventions for individuals. While SAR acceptance has been explored in general healthcare, limited evidence exists regarding its usability and acceptance among individuals with affective disorders. This study aimed to assess the usability and acceptance of "Pepper" as an adjunct motivational technology in combination with tablet-based cognitive training, compared to tablet training alone, among inpatients with depression, focusing on sex-specific differences. A randomized controlled trial was conducted between June and October 2024 with 32 inpatients diagnosed with depression. Participants were randomly assigned to one of two groups: the SAR group, which used "Pepper" together with the cognitive training app "Multimodal Activation" in combination with motivational feedback, and the control group, which used the same app on a tablet without motivational feedback. Each participant completed two cognitive training sessions of approximately 10 to 20 minutes. Standardised questionnaires, namely, *Technology Acceptance Model* (TAM), *System Usability Scale* (SUS), and *Technology Usage Inventory* (TUI), were administered after the second session. To analyze group and sex differences, analyses of co-variance were used for single-scale measures (SUS, TUI Overall),

10.3389/fpsyt.2025.1661730

and multiple analyses of co-variances were used for instruments with multiple subscales (TAM, TUI). A significant group difference was found in the SUS score, favoring the SAR group. The SAR group also scored higher in the TUI *User-Friendliness*, while the tablet group showed higher scores in *Accessibility*. Regarding sex differences, female participants scored higher on the TUI *Overall* and *Scepticism* subscales than male participants. These findings suggest that SAR-supported cognitive training is a viable and accepted tool for supporting cognitive training in psychiatric care. High usability and positive acceptance ratings indicate its potential to complement conventional therapies. The observed sex-specific differences underline the relevance of tailored robotic interventions. Limitations of this study include the small sample size and short intervention period. Further studies are warranted to validate these findings and examine the ethical considerations and human–robot interaction dynamics in psychiatric settings.

KEYWORDS

acceptance, usability, social assistive robotics (SAR), depression, cognitive training, sex differences, randomized controlled trial

1 Introduction

Depression is a widespread mental health disorder that affects over 264 million individuals worldwide (1). Treating depression often requires a multidimensional approach, including pharmacological treatment, psychotherapy/psychosocial interventions, and social support (2). In this context, socially assistive robots (SARs) could represent an innovative complement to traditional therapies and interventions. SARs are designed to interact with humans and assist them in various domains (3). The integration of robotic technology into healthcare has gained significant importance in recent years, particularly in mental health (4). SARs, such as the humanoid robot "Pepper", offer promising opportunities to support patients with mental health conditions, particularly within clinical and therapeutic settings. By enhancing social interaction during therapy sessions, promoting cognitive stimulation through interactive exercises, and fostering emotional well-being in individuals with depression and anxiety disorders, SARs can contribute significantly to mental healthcare. These effects are facilitated by socially assistive behaviors, including empathetic verbal responses, adaptive nonverbal communication (e.g. gestures, gaze, or posture), and the ability to tailor interactions based on the user's emotional state (5–7). However, the usability and acceptance of this technology among psychiatric patients, particularly those with depression, remain underexplored, and it is still unclear which patient groups are particularly responsive to SARs interventions, and which may find them less suitable.

Compared to other digital intervention platforms such as tablets or smartphones, SARs offer distinct advantages in promoting engagement and adherence in mental health interventions. Their embodied, human-like form enables multimodal communication, combining speech, gestures, and facial expressions, which fosters a stronger sense of social presence and emotional connection with users (3, 5, 6). Previous studies have shown that embodied agents can enhance motivation and therapeutic alliance more effectively than screen-based devices, particularly among individuals with affective or cognitive impairments (8–10).

While tablet-based cognitive training provides valuable accessibility and familiarity, it remains limited to two-dimensional interaction and lacks reciprocal social feedback. In contrast, humanoid SARs can actively perceive and respond to users' emotions, deliver verbal encouragement, and adapt their behavior in real time. These features make them particularly suitable for psychiatric rehabilitation, where emotional engagement and motivational support are critical for therapy adherence (11, 12).

The usability and acceptance of SARs are essential factors for their successful integration into clinical settings (13). Attitudes towards robots are influenced by cultural, demographic, and individual factors (14). High usability is crucial for both patients and healthcare providers because it facilitates effective use and encourages acceptance (15). This is especially important for patients with depression, where cognitive impairments and emotional factors may pose additional challenges to usability (16). Acceptance refers to patients' willingness to interact with the SAR and incorporate it into their treatment, whereas usability describes the robot's user-friendliness and effectiveness in practical applications (3). These aspects are particularly significant in healthcare, where perceived usefulness, user-friendliness, and trust in technology play pivotal roles (13). Research has consistently shown that the overall usability and acceptance of robots in healthcare are positive (5, 17–20). The usability and acceptance of SARs vary among different patient groups. For example, Pu et al. (21) found that older adults generally have a positive attitude towards SARs, particularly when these robots are perceived as supportive rather than replacements for human interaction. In

psychiatric care, Huijnen et al. (22) observed that patients with autism spectrum disorders benefit from interacting with SARs, indicating the potential usability and acceptance of SARs within this group. Broadbent et al. (3) emphasized that aspects of user-friendliness, such as intuitive interaction design, adaptability to user needs, and smooth navigation with minimal technical errors, are critical determinants of acceptance and willingness to engage with healthcare robots among older adults. This underscores the importance of designing intuitive and accessible human-robot interactions to ensure the successful adoption of SAR in diverse clinical settings.

Furthermore, baseline clinical factors such as depression severity may influence patients' engagement with digital or robot-assisted interventions. Higher symptom severity has been associated with reduced motivation, concentration, and psychomotor activity, which can limit active participation and perceived usability (23, 24). Likewise, pharmacological treatment may affect cognitive processing speed or emotional responsiveness, potentially moderating interaction quality and engagement. Considering these aspects is essential for understanding variability in patient responses to SAR-supported interventions (25).

Sex-specific differences in the acceptance and use of technology are also important aspects that must be considered in SAR research, particularly in psychiatric settings. Previous studies suggest that men and women may exhibit different preferences and attitudes towards robots, with some findings indicating higher acceptance of caregiving robots among women (26–28). These differences may significantly influence the perception and use of robots in clinical contexts. Despite the growing interest in SAR in healthcare, there is a lack of studies focusing on the usability and acceptance of such technologies among psychiatric inpatients with depression. Existing research has primarily examined older adults, individuals with dementia, and general medical populations (13, 14). Thus, there remains a need to explore how psychiatric patients, particularly those with depression, experience and evaluate SAR, while also accounting for sex-specific perspectives to inform the development of more tailored and sensitive robotic interventions (26).

This study aims to address recent research gaps by investigating the usability and acceptance of the SAR "Pepper" when used as an adjunct motivational technology in combination with tablet-based cognitive training, compared with tablet-based training alone without motivational feedback, among inpatients with depression in psychiatric care. Specifically, this study investigated how inpatients with depression in psychiatric care evaluate the usability and acceptance of a SAR-enhanced intervention, while additionally examining sex-specific differences in patients' evaluation of the intervention. For this purpose, we hypothesized that patients with depression who complete cognitive training sessions using the SAR "Pepper" with motivational feedback would report higher 1) usability and 2) acceptance than those receiving the same training via tablet, with significant sex-specific differences in their evaluations.

2 Methods and materials

2.1 Design, setting and sample

This pilot study was designed and conducted as a randomized controlled parallel two-arm trial (RCT) following the recommendations of the CONSORT 2025 Statement (29) to ensure transparency and methodological rigor in reporting. The RCT was carried out between June and October 2024 at the Clinical Division of Psychiatry and Psychotherapeutic Medicine at the Medical University of Graz, Austria. The sample consisted of 32 inpatients with depression. The diagnosis of depression was made by the treating clinicians according to the ICD-10 criteria. To evaluate the severity of the current depressive symptomatology the Montgomery-Åsberg Depression Rating Scale (MADRS) (30, 31) and the Beck Depression Inventory – Revised (BDI-II) (32) were used. The MADRS is an observer-rated assessment tool that evaluates depressive symptoms over the past week. It consists of ten items rated on a 7-point Likert scale, with total scores ranging from 0 to 60. Scores between 0 and 6 indicate no depression, whereas scores above 34 suggest severe depression (30, 31). In contrast, the BDI-II is a self-report questionnaire in which patients assess their feelings during the past week. It comprises 21 items rated on a 4-point Likert scale, yielding a maximum score of 64 points. A score of 0–8 points indicates no signs of depression, 9–13 points suggests minimal depression, 14–19 points indicates mild depression, 20–28 points indicates moderate depression, and scores of 29 or higher indicate severe depression (32). All assessments were conducted by trained staff members in a designated quiet room that was specifically prepared for this purpose. Both tests were used solely to describe the basic characteristics of the sample and were not considered as outcome variables.

2.2 Sample size

The sample size calculation was originally based on a repeated-measures ANCOVA design with two measurement points ($\eta^2 = 0.3$, power = .90, $\alpha = .05$), yielding a target of N = 32 participants (33). This approach was chosen because the broader project "AMIGA" (34, 35) included multiple pre–post measures, which are not part of the current publication. Although the present analysis focuses on post-intervention questionnaire data (SUS, TAM, TUI), the same total sample size was retained, as it meets the requirements for parametric testing based on the central limit theorem, which is generally satisfied for sample sizes of approximately 30 or more. "AMIGA" was conceptualized as a pilot study intended to generate initial evidence on usability and acceptance of SAR-supported cognitive training in depression. Consequently, the findings should be interpreted as exploratory and hypothesis-generating rather than confirmatory. The calculation was conducted using G*Power software (version 3.1.9.6) (36, 37).

10.3389/fpsyt.2025.1661730

2.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria for the participants were as follows. Participants eligible for inclusion were men and women aged ≥ 18 years who met the ICD-10 criteria for moderate or severe depression. They had to be able to speak and understand German, give written informed consent, and had no physical limitations that would prevent the use of the SAR "Pepper".

The exclusion criteria were refusal to participate, presence of psychotic disorders, organic brain diseases or dementia, substance-induced disorders, or acute suicidality.

2.4 Randomization and blinding

After the participants agreed to enroll, allocation to the intervention or control group was carried out through randomization using an online randomizer (https://www.studyrandomizer.com) (38), ensuring balanced randomization with respect to 1) sex, 2) depression severity according to the ICD-10 diagnosis at admission, and 3) age. Blinding was not performed in this study because the interventions were apparent and could not be concealed.

2.5 Study course

First, the participants were assigned to either an intervention group (SAR "Pepper" with motivational feedback) or a control group (tablet without motivational feedback). Over five days, participants in both groups had the opportunity to engage in cognitive training using the "Multimodal Activation" (MMA) application twice. In the intervention group, SAR "Pepper" was used to accompany and monitor the participants during training. After the participants completed the three tasks in cognitive training, SAR "Pepper" provided motivational feedback to encourage their progress. In contrast, the control group received the same cognitive training exercises, but these were conducted exclusively using the "MMA" app on an Android tablet (see Figures 1, 2). Apart from this difference, the procedures were identical in both groups. The primary distinctions between the intervention and control conditions were the mode of delivery (SAR "Pepper" versus tablet) and the presence or absence of motivational feedback during training (no motivational feedback in the tablet group).

The first training session began with the participants completing the baseline questionnaires (t_0) via LimeSurvey (https://www.limesurvey.org) (39). They then participated in the first cognitive training session. In the second session, they again completed a cognitive training session and filled out the post-intervention questionnaires (t_1) using LimeSurvey. Subsequently, the participants completed the study. Each cognitive training lasted approximately 10 to 20 minutes.

2.6 SAR "Pepper"

SAR "Pepper" is a 120 cm tall and 28 kg heavy humanoid robot. SAR "Pepper" is capable of moving its arms, hands, and head and can analyze and respond to human emotions through body posture, speech, facial expressions, and voice tone. Equipped with microphones, SAR "Pepper" can identify and turn towards the person speaking (40–42). SAR "Pepper" fosters friendly and effective human-robot interaction to support therapy, rehabilitation, or learning, often enhancing motivation and quality of life (43).

2.7 "MMA"-app and cognitive training

The "MMA" (m_ultim_odal a_ctivation) application (19, 44) consists of a tablet-based front end and a back end installed on a central server. Both components were developed using Microsoft's ".NET Framework". The front end was created using Xamarin, a cross-platform framework by Microsoft that enables mobile application development in ".NET". The application was specifically designed for Android and optimized for tablet displays. Additionally, an application programming interface was implemented to facilitate user management and data exchange, including authentication, version updates, and the transmission of exercise results. A structured query language database was developed to store the user data and the received results. Furthermore, a content management system was implemented, allowing the creation and categorization of various types of exercises into thematic units (45, 46).

Cognitive training consisted of three different exercise blocks, each comprising six tasks. Consequently, a total of 18 tasks were completed per training session. The difficulty level was adjusted after the completion of each exercise block and progressively increased with each step. The first exercise block was conducted at difficulty level 2 out of 4, the second block at difficulty level 3 out of 4, and the third block at the highest difficulty level, 4 out of 4.

Each training session included only three different cognitive tasks selected from distinct cognitive domains. In the first training session, participants were required to complete tasks such as finding picture pairs (domain: memory), solving arithmetic problems (domain: working memory), and filling in missing words in a text (domain: semantic memory). In the second training session, the tasks consisted of solving arithmetic problems (domain: working memory), completing puzzles (domain: visuospatial memory), and organizing a sequence of events in the correct order (domain: seriality and long-term memory).

The selection of cognitive tasks was based on established domains known to be affected in cognitive training, including memory, attention, executive functioning, and verbal fluency. Tasks were derived from the validated MMA training framework, which has previously been used in studies with neurological and psychiatric populations (19, 44–46). The difficulty levels were designed to

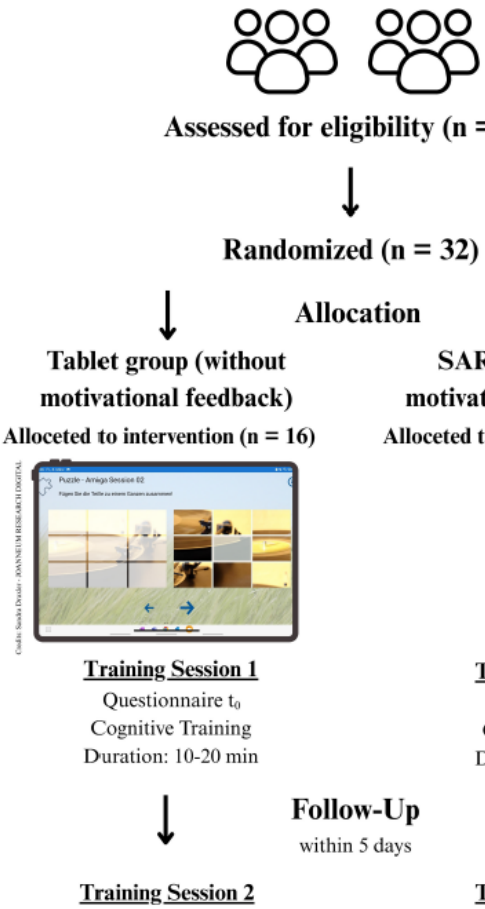

FIGURE 1
Study course and patient flow diagram according to Hopewell et al. (29).

progressively increase across sessions to sustain motivation and allow participants to experience a sense of achievement, while avoiding cognitive overload. This adaptive structure follows cognitive rehabilitation principles suggesting that gradual task escalation enhances engagement and supports learning in individuals with cognitive and affective impairments (47).

2.8 Questionnaires

2.8.1 System usability scale

The System Usability Scale (SUS) is a questionnaire designed to evaluate technology usability. The advantage of the SUS lies in its technology-independent nature and brevity, making it quick and

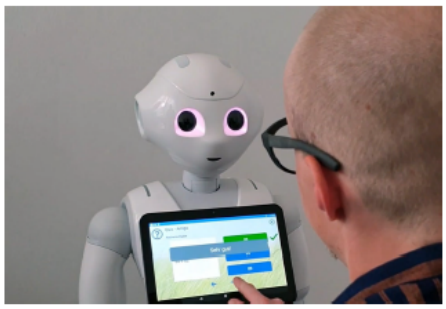

FIGURE 2
Participant interacts with the tablet-supported MMA application on the SAR "Pepper" and completes the cognitive task (solving arithmetic problems) (The participant provided written informed consent for this photograph to be taken and for its publication in the article; Credits: Sandra Draxler – JOANNEUM RESEARCH DIGITAL).

easy to complete. It consists of 10 questions rated on a five-point Likert scale (0 = strongly disagree to 4 = strongly agree). The items (2, 4, 6, 8, 10) need to be reverse-scored before further calculations, as they are negatively worded. Subsequently, the sum of all 10 items was calculated and multiplied by 2.5. This step yielded a total score (SUS score). A total score of 68 or higher indicates good usability (48–50). In the literature, the SUS is described as a reliable scale (ω = .89 and Cronbach's alpha α = .86) that effectively detects differences compared to other commercially available questionnaires, even with small sample sizes (51). Its validity, in this case, perceived user-friendliness, was also rated highly (52).

2.8.2 Technology usage inventory

The Technology Usage Inventory (TUI) evaluates the psychological and technology-specific factors that influence technology adoption. The TUI instrument consists of 30 items distributed across eight subscales: Curiosity, Anxiety, Interest, User-Friendliness, Usefulness, Scepticism, and Accessibility. The *Immersion* subscale, although part of the standard TUI, was intentionally excluded from this study, as immersion was not applicable to either the robot or tablet condition. This modification aligns with the TUI manual guidelines. Each item was rated on a seven-point Likert scale (1 = "Strongly Disagree" to 7 = "Strongly Agree"). Subscales with three items (Accessibility and User-Friendliness) yielded scores ranging from 3 to 21 points, while subscales with four items (Anxiety, Curiosity, Interest, Usefulness, and Scepticism) ranged from 4 to 28 points, resulting in a total possible TUI score between 26 and 182 points. Additionally, the Intention to Use scale, comprising three items, was included, allowing participants to rate their intentions on a scale from 0 to 100 per item, for a total score ranging from 0 to 300. The TUI has demonstrated robust psychometric properties, with internal consistencies (Cronbach's alpha) ranging from.70 to.89 (53).

2.8.3 Technology acceptance model

The Technology Acceptance Model (TAM) by Davis (54) represents a central approach in acceptance research and serves as a foundation for many subsequent models (55). This theory aims to explain the individual usage behavior of new technologies and their influencing factors, with the goal of predicting technology use. This model is based on two significant influencing factors: 'perceived usefulness' and 'perceived ease of use' of the technology. Perceived usefulness refers to the extent to which individuals believe that using a particular system enhances their performance (54). Therefore, a high level of perceived usefulness is believed to positively influence users' performance. Perceived ease of use describes the extent to which individuals assume that using a system requires little effort (54). Individuals evaluate a technology with a high perceived ease of use as simple to operate (54). Each scale consisted of six questions. The two subscales Perceived Usefulness and Perceived Ease of Use demonstrated excellent internal consistency, with Cronbach's alphas of.98 and.94, respectively (54). Additionally, two further questions were included to assess the intention to use the app. The questionnaire consisted of 14 items, which were answered using a seven-point Likert scale (1 = "Strongly Disagree" to 7 = "Strongly agree").

2.9 Analysis

All statistical analyses were performed using IBM SPSS Statistics software (version 30) (56). Demographic characteristics were analyzed using descriptive statistics, including means (*M*) and standard deviations (*SD*). The usability and acceptance scales were evaluated exclusively at the second measurement point (t_1). To examine group and sex differences in the questionnaire outcomes, different statistical approaches were applied based on the structure of the respective instruments. For questionnaires consisting of a single overall scale (i.e., SUS and the overall score of the TUI), a one-way analysis of co-variance (ANCOVA) was performed. For instruments comprising multiple subscales (i.e., TAM and the subscales of the TUI), a multivariate analysis of co-variance (MANCOVA) was conducted. Statistical significance was set at p <.05 (two-tailed hypothesis testing) for all analyses. Prior to conducting ANCOVA and MANCOVA, all statistical assumptions were tested and met. To account for potential baseline imbalances between group and sex, ANCOVAs and MANCOVAs were adjusted for baseline depression severity (BDI-II t_0 and MADRS t_0 scores). Adjusted effect sizes (partial η^2) and 95% confidence intervals (*CI*s) are reported. Bonferroni corrections were applied to control for multiple comparisons in both the ANCOVA and MANCOVA analyses.

TABLE 1 Demographic and clinical characteristics of the sample.

Variable	SAR group (with motivational feedback)	Tablet group (without motivational feedback)	Total
Female (*n*)	8	8	16
Male (*n*)	8	8	16
Age *(M, SD)*	37.69 (14.59)	38.56 (13.91)	38.12 (14.03)
MADRS *(M, SD)*	32.06 (4.37)	29.19 (5.28)	30.63 (4.99)
BDI-II *(M, SD)*	28.25 (9.61)	22.63 (6.51)	25.44 (8.57)

BDI-II, Beck Depression Inventory; M, Mean; MADRS, Montgomery Åsberg Depression Rating Scale; SAR, Socially Assistive Robot; SD, Standard Deviation.

3 Results

3.1 Characteristics of the participants

The study included 32 inpatients with moderate to high depression symptom severity according to ICD-10 criteria. The mean age of participants was 38.12 years (*SD* = 14.03; range = 19–65), with an equal distribution of women (*n* = 16) and men (*n* = 16) across the SAR group (with motivational feedback) and tablet group (without motivational feedback). Overall, the participants presented with moderate depressive symptomatology, as indicated by mean MADRS scores of 30.63 (± 4.99) and BDI-II scores of 25.44 (± 8.57). The groups were comparable in age and sex composition, while the SAR group (with motivational feedback) showed slightly higher baseline depression severity than the tablet group (without motivational feedback) (see Table 1). No participants dropped out during the study period, and all completed both training sessions and questionnaires. Data were checked for completeness before analysis; no cases with missing values were identified. Therefore, all 32 participants were included in the final analyses. Mean scores (*M*), standard deviations and (*SD*) and 95% confidence intervals of the SUS, the TUI and the TAM are presented in Table 2.

3.2 System usability scale

Bonferroni-corrected *post-hoc* analysis revealed a significant difference between SUS scores of the SAR group (with motivational feedback) and the tablet group (without motivational feedback (p <.001, M_{Dif} = 15.317, 95%-CI[6.938, 23.696]), after adjusting for baseline depression scores (BDI-II t_0 and MADRS t_0). No statistically significant results were found for sex comparisons (see Table 3).

3.3 Technology usage inventory

Bonferroni-corrected *post-hoc* analysis revealed a significant difference between TUI overall scores of females and males (p = .033, M_{Dif} = 11.479, 95%-CI [.974, 21.983]), after adjusting for baseline Depression scores (BDI-II t_0 and MADRS t_0) (see Table 3).

After adjusting for baseline depression scores (BDI-II t_0 and MADRS t_0), a multivariate analysis of covariance (MANCOVA) revealed significant group differences for two subscales of the TUI.

Participants in the SAR group (with motivational feedback) reported significantly higher scores in the *User-Friendliness* subscale compared to the Tablet group (without motivational feedback), F (*1*, 28) = 12.18, p = .002, η^2_p = .303, 95% CI [1.81, 6.95]. In contrast, the tablet group showed significantly higher scores in the subscale *Accessibility* than the SAR group, F (*1*, 28) = 8.07, p = .008, η^2_p = .224, 95% CI [−6.70, −1.09].

A further significant effect was found for sex in the subscale *Scepticism*, with female participants reporting higher scores than male participants, F(1, 28) = 4.34, p = .046, η^2_p = .134, 95% CI [0.06, 6.90].

3.4 Technology acceptance model

After adjusting for baseline depression scores (BDI-II t_0 and MADRS t_0), a multivariate analysis of covariance (MANCOVA) revealed no statistically significant results for all subscale for group and sex comparisons (see Table 4).

4 Discussion

This study aimed to evaluate the usability and acceptance of the SAR "Pepper" when implemented as a motivational adjunct to tablet-based cognitive training, compared to cognitive training delivered without robotic support, in inpatients with depression, with a particular focus on sex-specific differences. These findings provide valuable insights into the integration of robotic technologies into psychiatric care. Given the increasing interest in technology-based interventions for mental health care, understanding user preferences and technological acceptance is crucial for the successful implementation of such interventions (8).

Regarding technology usability and acceptance, the results revealed several significant findings. Usability ratings measured via the SUS and user-friendliness measured via TUI User-Friendliness favored the SAR group (with motivational feedback), with significantly higher scores compared to the tablet group (without motivational feedback). In contrast the tablet group (with motivational feedback) showed significantly higher scores in the TUI subscale Accessibility. Additionally, sex-related differences were observed, with female participants demonstrating significantly higher TUI Overall and Scepticism subscale scores than males.

TABLE 2 Descriptive statistics of system usability scale, technology usage inventory and technology acceptance modell between SAR and tablet group and sex.

Variables	SAR group (with motivational feedback) (n = 16)				Tablet group (without motivational feedback) (n = 16)			
	M	*SD*	95% CI		*M*	*SD*	95% CI	
			Lower bound	*Upper bound*			*Lower bound*	*Upper bound*
SUS	77.50	10.165	72.08	82.93	61.72	9.904	56.44	67.00
TUI OVERALL	98.94	13.051	91.98	105.89	102.19	16.710	93.28	111.09
TUI INT	19.25	4.973	16.60	21.90	16.06	5.221	13.28	18.84
TUI USE	16.00	7.659	11.92	20.08	17.88	4.500	15.48	20.27
TUI SCE	14.13	4.689	11.63	16.62	10.13	4.530	7.71	12.54
TUI ANX	13.25	4.669	10.76	15.74	9.88	5.772	6.80	12.95
TUI UF	17.63	3.594	15.71	19.54	14.56	3.098	12.91	16.21
TUI ACC	7.56	3.054	5.93	9.19	10.31	4.029	8.17	12.46
TUI CUR	18.50	5.514	15.56	21.44	16.00	5.797	12.91	19.09
TUI ITU	206.38	37.137	186.59	226.16	185.00	77.562	143.67	226.33
TAM PU	30.56	3.949	28.46	32.67	32.63	3.914	30.54	34.71
TAM PEU	34.19	3.582	32.28	36.10	30.88	4.544	28.45	33.30
TAM ITU	11.88	1.586	11.03	12.72	11.00	1.592	10.15	11.85
Variables	Female (n = 16)				Male (n = 16)			
	M	*SD*	95% CI		*M*	*SD*	95% CI	
			Lower bound	*Upper bound*			*Lower bound*	*Upper bound*
SUS	69.69	13.130	62.69	76.68	69.53	12.722	62.75	76.31
TUI OVERALL	106.38	16.272	97,70	115.05	94.75	10.878	88.95	100.55
TUI INT	17.56	5.278	14.75	20.38	17.75	5.434	14.85	20.65
TUI USE	18.31	6.750	14.72	21.91	15.56	5.585	12.59	18.54
TUI SCE	13.69	5.056	10.99	16.38	10.56	4.501	8.16	12.96
TUI ANX	12.06	5.531	9.12	15.01	11.06	5.482	8.14	13.98
TUI UF	16.63	3.500	14.76	18.49	15.56	3.829	13.52	17.60
TUI ACC	9.44	4.589	6.99	11.88	8.44	2.828	6.93	9.94
TUI CUR	18.69	5.724	15.64	21.74	15.81	5.492	12.89	18.89
TUI ITU	210.94	56.600	180.78	241.10	180.44	62.762	146.99	213.88
TAM PU	32.81	4.151	30.60	35.02	30.38	3.575	28.47	32.28
TAM PEU	33.25	4.583	30.81	35.69	31.81	4.151	29.60	34.02
TAM ITU	11.69	1.401	10.94	12.43	11.19	1.834	10.21	12.16

M and *SD* are used to represent mean and standard deviation including the 95% confidence interval for each item.
ACC, Accessibility; ANX, Anxiety; CI, confidence interval; CUR, Curiosity; INT, Interest; ITU, Intention to Use; M, Mean; PEU, Perceived Ease of Use; PU, Perceived Usefulness; SAR, Socially Assistive Robot; SCE, Scepticism; SD, Standard deviation; Sig, Significance; SUS, System Usability Scale; TAM, Technology Acceptance Model; TUI, Technology Usage Inventory; UF, User-Friendliness; USE, Usefulness.

These findings highlight that while the robot-assisted intervention was perceived as usable and user-friendly, it was also met with higher skepticism. This finding indicates that higher usability does not necessarily translate into greater overall acceptance. The SAR condition appears to evoke ambivalent reactions, as participants appreciated the interactive and social features while simultaneously expressing reservations regarding the SAR interaction.

SAR "Pepper's" usability was also rated positively, indicating a user-friendly design for human-robot interaction for patients with

TABLE 3 Comparison of system usability scale and technology usage inventory overall between SAR and tablet group and sex after adjusting for baseline depression scores (BDI-II t_0 and MADRS t_0).

Dependent variable	(I)	(J)	Mean difference (I-J)	*F*	Sig. [a]	η^2_p	95% CI for difference [a]	
							Lower bound	Upper bound
SUS	SAR [b]	Tablet [c]	15.317	14.022	**<.001*** **	.334	6.938	23.696
TUI OVERALL	SAR [b]	Tablet [c]	-4.635	.582	.452	.020	-17.085	7.815
SUS	Female	Male	-.104	-001	.982	.000	-9.411	9.204
TUI OVERALL	Female	Male	11.479	5.011	**.033***	.152	.974	21.983

ANCOVA. *N* = 32.
η^2p, Partial Eta Squared; SAR, Socially Assistive Robot; Sig, Significance; SUS, System Usability Scale; TUI, Technology Usage Inventory.
* p <.05. ** p <.01. *** p <.001.
[a]Adjustment for multiple comparisons: Bonferroni.
[b]with motivational feedback.
[c]without motivational feedback.
The values shown in bold indicate statistically significant results. Depending on the level of significance achieved, they are marked with *, **, or ***.

TABLE 4 Comparison of the technology usage inventory and the technology acceptance model between SAR and tablet group and sex after adjusting for baseline depression scores (BDI-II t_0 and MADRS t_0).

Dependent variable	(I)	(J)	Mean difference (I-J)	*F*	Sig. [a]	η^2_p	95% CI for difference [a]	
							Lower bound	Upper bound
TUI INT	SAR [b]	Tablet [c]	3.744	3.315	.079	.106	-.472	8.019
TUI USE	SAR [b]	Tablet [c]	-1.997	.604	.444	.021	-7.264	3.269
TUI SCE	SAR [b]	Tablet [c]	3.583	3.651	.066	.115	-.258	7.424
TUI ANX	SAR [b]	Tablet [c]	2.944	2.150	.154	.071	-1.169	7.056
TUI UF	SAR [b]	Tablet [c]	4.381	12.175	**.002****	.303	1.809	6.953
TUI ACC	SAR [b]	Tablet [c]	-3.893	8.069	**.008****	.224	-6.698	-1.085
TUI CUR	SAR [b]	Tablet [c]	3.701	2.683	.113	.087	-.927	8.330
TUI ITU	SAR [b]	Tablet [c]	34.995	2.315	.139	.076	-12.121	82.111
TAM PU	SAR [b]	Tablet [c]	-2.906	3.522	.071	.112	-82.111	12.121
TAM PEU	SAR [b]	Tablet [c]	2.568	2.512	.124	.082	-.751	5.888
TAM ITU	SAR [b]	Tablet [c]	.803	1.547	.224	.052	-.520	2.126
TUI INT	Female	Male	-.143	.005	.943	.000	-4186	3.900
TUI USE	Female	Male	2.714	1.413	.245	.048	-1.963	7.391
TUI SCE	Female	Male	3.480	4.340	**.046***	.134	.058	6.902
TUI ANX	Female	Male	1.283	.475	.496	.017	-2.528	5.093
TUI UF	Female	Male	.959	.510	.481	.018	-1.791	3.708
TUI ACC	Female	Male	1.110	.643	.429	.022	-1.726	3.946
TUI CUR	Female	Male	2.792	1.830	.187	.061	-1.435	7.019
TUI ITU	Female	Male	29.271	1.974	.171	.066	-13.399	71.941
TAM PU	Female	Male	2.549	3.320	.079	.106	-.317	5.414
TAM PEU	Female	Male	1.808	1.483	.233	.050	-1.233	4.849
TAM ITU	Female	Male	.570	.941	.340	.033	-.633	1.774

MANCOVA. *N* = 32.
ACC, Accessibility; ANX, Anxiety; CUR, Curiosity; INT, Interest; ITU, Intention to Use. η^2p, Partial Eta Squared; PEU, Perceived Ease of Use; PU, Perceived Usefulness; SAR, Socially Assistive Robot; Sig, Significance; SCE, Scepticism; TAM, Technology Acceptance Model; TUI, Technology Usage Inventory; UF, User-Friendliness; USE, Usefulness.
*p <.05. **p <.01. ***p <.001.
[a]Adjustment for multiple comparisons: Bonferroni.
[b]with motivational feedback.
[c]without motivational feedback.
The values shown in bold indicate statistically significant results. Depending on the level of significance achieved, they are marked with *, **, or ***.

depression. These results align with previous research, indicating that SARs tend to be rated as highly engaging due to their interactive nature and human-like characteristics (8, 57). For example, the robot 'Ryan', which delivers internet-based cognitive behavioral therapy, has been shown to improve mood scores and engagement in older adults with depression (9). This underscores the importance of intuitive usability for the successful integration of robotic technology into clinical environments. Similar findings were reported by (11), who demonstrated that high usability improved SAR acceptance and utilization among older adults. Notably, patients with depression, who often face cognitive impairments (16), reported no significant difficulties in operating the robot. The user-friendly approach of SAR "Pepper" may be attributed to its perception as a supportive complement to traditional therapy rather than a replacement for human interaction (5). This interpretation is supported by Göransson et al. (58), who found that patients evaluated SARs more positively when they were presented as an addition rather than a substitute for human contact. Broadbent (3) emphasized the importance of intuitive and user-friendly design in overcoming barriers to technology use and facilitating its integration into clinical settings.

Qualitative feedback collected informally after the sessions supported these findings: several participants in the SAR group described SAR "Pepper" as "interesting but somewhat unusual at first," and some expressed uncertainty about "how natural" the interaction felt compared with human contact. Others reported technical aspects such as delayed speech recognition or limited gesture responsiveness as minor obstacles to immersion. Similar concerns have been described in previous research, where users reported feelings of unfamiliarity, artificiality, or reduced authenticity in early stages of human–robot interaction (59, 60).

Furthermore, these findings align with those of Bishop et al. (17), who documented a generally positive attitude towards healthcare robots. Da Rocha Ferreira et al. (5) also highlighted the importance of clear communication regarding the added value of SAR in driving their acceptance. The results also showed that interacting with "Pepper" was perceived as enjoyable and motivating. This is particularly important because motivation and engagement play central roles in treating depression. The positive perception of usability underscores the importance of well-thought-out human-robot interactions to foster the acceptance and practical use of SAR.

Despite these advantages, skepticism toward SARs remains a notable barrier to acceptance. In our study, females reported significantly higher skepticism scores than men. This finding aligns with research suggesting that while SARs may be initially engaging, long-term usability and trust issues may arise, particularly among individuals with higher levels of anxiety or resistance to new technologies (61). Similar concerns were raised in studies evaluating the robot 'Paro', where some participants expressed reluctance to engage with a robotic intervention, perceiving it as artificial or intrusive (10, 62).

The significantly higher accessibility scores reported by participants in the tablet group (without motivational feedback) compared to those in the SAR group (with motivational feedback) can be attributed to a combination of cost- and usability-related factors. Tablets are generally more affordable and readily available, making them more accessible for a broader user base (63, 64). In contrast, SARs involve high upfront costs because of their complex design, restricting their use primarily to high-income settings (65). From a usability perspective, tablets benefit from familiar touchscreen interfaces and intuitive navigation, although certain design aspects, such as small screen size or complex applications, can potentially hinder their effectiveness for users with physical or cognitive limitations (66). Although SARs are specifically designed for ease of use, their technical limitations, including the need for structured training, may negatively impact perceived accessibility among older or mentally ill users (67). Overall, these findings underscore that the tablet's combination of economic feasibility and relative usability likely contributed to its higher accessibility ratings in the current study.

A central aspect of this study was the exploration of sex-specific differences in the usability and acceptance of SAR "Pepper". The results revealed that women tended to have a higher overall usability of SARs than men. This observation is consistent with the findings of Kuo et al. (26), who examined age- and sex-specific differences in the perception and acceptance of healthcare robots. Moradbekhti et al. (27) also found that women rated SARs more positively, particularly regarding its usefulness and user-friendliness. The reasons for these sex-specific differences are complex. Ringwald et al. (28) argue that such differences may be attributed to distinct socialization experiences and role expectations. Women, often characterized by stronger caregiving orientations and social sensitivity, may be more receptive to using supportive technologies. Additionally, they may place greater emphasis on the emotional and social aspects of interacting with SARs than men. This suggests that the relationship between sex and SAR acceptance is complex and potentially context-dependent. The sex-specific findings of this study highlight the necessity of incorporating sex-sensitive approaches into the development and implementation of SARs. This could involve targeted adjustments in human-robot interaction or differentiated communication strategies that address the specific needs and expectations of different sexes.

Another critical aspect of SARs adoption is the role of sex-specific preferences. Recent research suggests that men and women may have different expectations of SARs, with women placing greater emphasis on emotional support and social interaction and men prioritizing functional aspects and usability (68). This may explain some of the variability in user engagement and skepticism observed in the present study. Furthermore, studies have highlighted that the personality adaptation of SARs can significantly influence user acceptance. For example, SARs with extroverted personalities tend to be more accepted by extroverted users, whereas introverted users may prefer calmer, less expressive robotic interactions (69, 70). These findings emphasize the importance of adaptive, user-tailored robot interactions, a concept that has been incorporated into SAR interventions, such as this project (35).

A recurring challenge in SAR-based interventions is the sustainability of engagement over time. While our study found a

LI

preference for the SAR over the tablet in terms of usability and user friendliness, it remains unclear whether this preference would persist over an extended period. Previous research indicates that SAR engagement often decreases after an initial novelty phase, leading to reduced long-term adherence (10, 62). A possible explanation for this phenomenon is the limited adaptability of SARs, which may lead users to perceive interactions as repetitive or lacking meaningful engagement over time (12).

Additionally, ethical concerns regarding privacy, data security, and autonomy have been raised in previous studies, particularly in relation to psychiatric patients (61). While SARs offer promising solutions for cognitive training and emotional support, their successful implementation in clinical practice requires addressing the above-mentioned barriers to trust and acceptance.

The positive results regarding the usability and acceptance of SAR "Pepper" among patients with depression highlight the promising opportunities for SARs use in psychiatric care. SARs can serve as a complementary tool in depression treatment, assisting with cognitive training exercises, providing information, and offering emotional support. Studies such as those by Chen et al. (71) have demonstrated that SAR-based intervention programs can positively affect cognitive function and mood in older adults in long-term care facilities. Similarly, Chen et al. (72) reported successful SAR interventions in children with autism spectrum disorders. These findings suggest that SARs could be versatile tools for treating depression.

5 Strengths and limitations

This study provides valuable insights into the usability and acceptance of SARs and tablet-based cognitive training in individuals with depression. A key strength of this study is the randomized group assignment, which minimizes selection bias and ensures comparability between the intervention groups. By using validated psychometric instruments, such as the MADRS and BDI-II, this study offers robust clinical assessments of depressive symptoms, increasing the reliability of the findings.

Another strength of this study is the use of multiple standardized measures to assess technology usability and acceptance. The inclusion of the SUS, the TUI, and the TAM allows for a comprehensive evaluation of the user experience.

Additionally, this study contributes to the growing body of research on SARs in mental health care by examining their clinical usability. Previous research has mainly focused on older adults or individuals with cognitive impairments (9, 57), whereas this study specifically addressed depression in an adult population, thereby expanding the applicability of SAR interventions to broader psychiatric settings.

Despite its promising findings and strengths, this study has several limitations that should be acknowledged. The main limitation concerns the asymmetry in feedback between the two groups. The SAR group included verbal motivational feedback of the SAR "Pepper", whereas the tablet condition did not provide any motivational feedback beyond task progression. This introduces an additional factor, that may have contributed to the higher usability and acceptance ratings in the SAR group. Consequently, the results should be interpreted as reflecting the combined effect of humanoid interaction and motivational feedback rather than the robotic embodiment alone. Future studies should equalize feedback across conditions to isolate the specific contribution of the SAR component.

The second limitations was the relatively small the sample size ($n = 32$), which may limit the generalizability of the findings. While the randomized design strengthens internal validity, the statistical power of the study remains restricted, particularly in detecting subtle effects or interaction effects between the variables. Future studies should aim to replicate these findings in larger and more diverse samples to enhance their external validity.

Another limitation relates to the limited intervention duration and training frequency of the cognitive training intervention. Participants completed only two training sessions, which may not have been sufficient to produce measurable improvements in cognitive performance. Prior studies have indicated that long-term engagement is crucial for the sustained benefits of SAR-based interventions, as engagement levels may decline after an initial novelty phase (10, 62). Abdi et al. (73) emphasized the importance of long-term studies to better understand the lasting impact and acceptance of SARs in healthcare settings. Future research should explore the effects of prolonged SAR interventions on cognitive and emotional outcomes.

The sex-specific differences observed in this study warrant further investigation in future studies. Intersectional aspects should also be considered to develop a nuanced understanding of the factors influencing SAR usability and its acceptance. Eyssel and Hegel (74) emphasized the need to incorporate cultural and socioeconomic factors into the analysis to obtain a holistic perspective on this issue.

This study focused exclusively on patients with depression. Future research should explore the use of SARs for other psychiatric conditions, such as anxiety disorders, bipolar disorders, and post-traumatic stress disorders. Studies by Huijnen et al. (22) and Rabbitt et al. (75) have shown promising results for the use of SARs among patients with autism spectrum disorders and other psychiatric diseases.

Beyond the methodological and conceptual limitations discussed above, two additional aspects warrant further attention. The first relates to the integration of SARs into existing treatment frameworks. While this study focused on the evaluation of usability and acceptance, it did not examine how SAR-based cognitive training interacts with standard therapeutic approaches. Therefore, future studies should investigate whether combining SAR interventions with conventional treatments can enhance therapeutic outcomes. As highlighted by Tapus et al. (76), such

integrative approaches are crucial for harnessing the full potential of SARs in psychiatric care.

Another important consideration is the ethical dimensions of SAR use in sensitive clinical environments. Although not the focus of this study, issues such as data privacy, informed consent, and unintended psychosocial consequences of human-robot interactions are highly relevant, particularly when working with vulnerable populations. Addressing these questions in future research is essential to ensure ethical implementation. Sharkey and Sharkey (77) emphasize the importance of proactively engaging with ethical challenges in the context of healthcare robotics.

6 Conclusion

The overall positive evaluation of SAR "Pepper" underscores the potential of SARs as a complementary tool in psychiatric care. The observed sex-specific differences highlight the need for tailored approaches to developing and implementing SAR technology in clinical environments.

The results of this study provide a solid foundation for further research and advancement of SAR technology in psychiatric care. They highlight the potential of SAR technology to enhance treatment quality and open new pathways for patient care. Simultaneously, they caution against an overreliance on SAR and stress the need to view them as complements, not replacements, to human interaction and traditional therapies.

Future research should investigate the long-term effects of SAR in psychiatric care, extend its applicability to other mental health conditions, and explore complex interactions between SAR technology and traditional treatment approaches. Only through continuous research and careful evaluation can we fully harness the potential of SAR in psychiatric care while ensuring their use remains ethical and patient centered.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Medical University of Graz (EK Nr. 35-450 ex 22/23). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

AH: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. IZ: Data curation, Investigation, Resources, Writing – review & editing. LS: Data curation, Resources, Writing – review & editing. IS: Data curation, Resources, Writing – review & editing. MSt: Data curation, Resources, Writing – review & editing. TS: Formal Analysis, Writing – review & editing. ML: Conceptualization, Funding acquisition, Project administration, Writing – review & editing. FF: Conceptualization, Funding acquisition, Project administration, Writing – review & editing. SG: Funding acquisition, Project administration, Writing – review & editing. ES: Resources, Writing – review & editing. MP: Conceptualization, Funding acquisition, Writing – review & editing, Data curation, Methodology. TO: Conceptualization, Funding acquisition, Writing – review & editing, Data curation, Methodology. SD: Conceptualization, Funding acquisition, Writing – review & editing, Methodology. MSc: Conceptualization, Funding acquisition, Writing – review & editing, Methodology. SR: Conceptualization, Funding acquisition, Project administration, Writing – review & editing, Methodology, Supervision. JZ: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. DS: Data curation, Writing – review & editing. AS: Data curation, Writing – review & editing. SS: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. MM: Project administration, Writing – review & editing, Conceptualization, Funding acquisition, Methodology, Supervision. LP: Conceptualization, Funding acquisition, Project administration, Writing – review & editing, Resources, Supervision, Methodology. ND: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing, Methodology. ER: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was partly funded by project AMIGA under grant No FO999892829 of the Austrian FFG femtech programme co-financed by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK). #AH will receive funding to cover the publication fee costs for this article from the Doctoral School "Lifestyle-related Diseases" at the Medical University of Graz.

Acknowledgments

The authors thank all the participants for their cooperation in this study. Additionally, we would like to express our gratitude to Humanizing Technologies GmbH, especially Dimitrios Prodromou, for their support in the technical implementation of the SAR "Pepper." This publication was prepared in partial

fulfilment of the requirements for the doctoral program in Medical Science (doctoral school "Lifestyle Related Diseases") at the Medical University of Graz, Austria.

Conflict of interest

Author MM was employed by Humanizing Technologies GmbH.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

1. World Health Organization (WHO). Monitoring mental health systems and services in the WHO European Region: Mental Health Atlas. *World Health Organization. Regional Office for Europe*. (2019). Available online at: https://iris.who.int/handle/10665/346404.

2. Malhi GS, Bell E, Bassett D, Boyce P, Bryant R, Hazell P, et al. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry*. (2021) 55:7–117. doi: 10.1177/0004867420979353

3. Broadbent E. Interactions with robots: the truths we reveal about ourselves. *Annu Rev Psychol*. (2017) 68:627–52. doi: 10.1146/annurev-psych-010416-043958

4. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of social robots in mental health and well-being research: systematic review. *J Med Internet Res*. (2019) 21:e13322. doi: 10.2196/13322

5. Da Rocha Ferreira AR, Silva MGMMDS, De Oliveira Correia M, De Almeida RS, Simões-Silva V, Trigueiro MJ. Use of Socially Assistive Robots in Mental Health: Barriers and Facilitators. In: Simões De Almeida R, Simões-Silva V, Trigueiro MJ, editors. *Handbook of Research on Advances in Digital Technologies to Promote Rehabilitation and Community Participation*. IGI Global Scientific Publishing (2023). p. 269–92. doi: 10.4018/978-1-6684-9251-2.ch013

6. Spitale M, Okamoto S, Gupta M, Xi H, Matarić MJ. Socially assistive robots as storytellers that elicit empathy. *ACM Trans Hum-Robot Intenact*. (2022) 11:1–29. doi: 10.1145/3538409

7. Yee LOH, Fun SS, Zin TS, Aung ZN, Yap KM, Teoh J. Socially assistive robots: A technological approach to emotional support. (2024). doi: 10.48550/ARXIV.2411.05122

8. Lee HR, Šabanović S, Chang W-L, Nagata S, Piatt J, Bennett C, et al. (2017). Steps toward participatory design of social robots: mutual learning with older adults with depression, in: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna Austria. pp. 244–53. doi: 10.1145/2909824.3020237

9. Dino F, Zandie R, Abdollahi H, Schoeder S, Mahoor MH. (2019). Delivering cognitive behavioral therapy using A conversational social robot, in: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China. pp. 2089–95. doi: 10.1109/IROS40897.2019.8968576

10. Randall N, Bennett CC, Šabanović S, Nagata S, Eldridge L, Collins S, et al. More than just friends: in-home use and design recommendations for sensing socially assistive robots (SARs) by older adults with depression. *Paladyn J Behav Robot*. (2019) 10:237–55. doi: 10.1515/pjbr-2019-0020

11. Khosla R, Nguyen K, Chu M-T. Human robot engagement and acceptability in residential aged care. *Int J Human–Computer Intenact*. (2017) 33:510–22. doi: 10.1080/10447318.2016.1275435

12. Guemghar I, Pires de Oliveira Padilha P, Abdel-Baki A, Jutras-Aswad D, Paquette J, Pomey M-P. Social robot interventions in mental health care and their outcomes, barriers, and facilitators: scoping review. *JMIR Ment Health*. (2022) 9:e36094. doi: 10.2196/36094

13. Rantanen T, Lehto P, Vuorinen P, Coco K. The adoption of care robots in home care—A survey on the attitudes of Finnish home care personnel. *J Clin Nurs*. (2018) 27:1846–59. doi: 10.1111/jocn.14355

14. Papadopoulos I, Wright S, Koulouglioti C, Ali S, Lazzarino R, Martín-García À, et al. Socially assistive robots in health and social care: Acceptance and cultural factors. Results from an exploratory international online survey. *Jpn J Nurs Sci*. (2023) 20:e12523. doi: 10.1111/jjns.12523

15. Moharana S, Panduro AE, Lee HR, Riek LD. (2019). Robots for joy, robots for sorrow: community based robot design for dementia caregivers, in: *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Daegu, Korea (South. pp. 458–67. doi: 10.1109/HRI.2019.8673206

16. Ryan M, Stahl BC. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *J Inf Commun Ethics Soc*. (2021) 19:61–86. doi: 10.1108/JICES-12-2019-0138

17. Bishop L, Van Maris A, Dogramadzi S, Zook N. Social robots: The influence of human and robot characteristics on acceptance. *Paladyn J Behav Robot*. (2019) 10:346–58. doi: 10.1515/pjbr-2019-0028

18. Paletta L, Brijacak I, Reiterer B, Pszeida M, Ganster H, Fuhrmann F, et al. (2019). Gaze-based human factors measurements for the evaluation of intuitive human-robot collaboration in real-time, in: *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Zaragoza, Spain. pp. 1528–31. doi: 10.1109/ETFA.2019.8869270

19. Paletta L, Schüssler S, Zuschnegg J, Steiner J, Pansy-Resch S, Lammer L, et al. AMIGO—A socially Assistive Robot for Coaching Multimodal Training of Persons with Dementia. In: Korn O, editor. *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*. Springer International Publishing, Cham (2019). p. 265–84. doi: 10.1007/978-3-030-17107-0_13

20. Zuschnegg J, Schoberer D, Häussl A, Herzog SA, Russegger S, Ploder K, et al. Effectiveness of computer-based interventions for community-dwelling people with cognitive decline: a systematic review with meta-analyses. *BMC Geriatr*. (2023) 23:229. doi: 10.1186/s12877-023-03941-y

21. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: A systematic review and meta-analysis of randomized controlled studies. *Gerontologist*. (2019) 59:e37–51. doi: 10.1093/geront/gny046

22. Huijnen CA, As Lexis M, De Witte LP. Robots as new tools in therapy and education for children with autism. *Int J Neurorehabilitation*. (2017) 04. doi: 10.4172/2376-0281.1000278

23. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*. (2021) 20:318–35. doi: 10.1002/wps.20883

24. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. *Curr Treat Options Psychiatry*. (2023) 10:119–35. doi: 10.1007/s40501-023-00297-3

25. Guerrera CS, Platania GA, Boccaccio FM, Sarti P, Varrasi S, Colliva C, et al. The dynamic interaction between symptoms and pharmacological treatment in patients with major depressive disorder: the role of network intervention analysis. *BMC Psychiatry*. (2023) 23:885. doi: 10.1186/s12888-023-05300-y

26. Kuo IH, Rabindran JM, Broadbent E, Lee YI, Kerse N, Stafford RMQ, et al. (2009). Age and gender factors in user acceptance of healthcare robots, in: *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, Japan. pp. 214–9. doi: 10.1109/ROMAN.2009.5326292

27. Moradbakhti L, Mara M, Castellano G, Winkle K. (Counter-)stereotypical gendering of robots in care: impact on needs satisfaction and gender role concepts in men and women users. *Int J Soc Robot*. (2023) 15:1769–90. doi: 10.1007/s12369-023-01033-w

LIV

28. Ringwald M, Theben P, Gerlinger K, Hedrich A, Klein B. How should your assistive robot look like? A scoping review on embodiment for assistive robots. *J Intell Robot Syst.* (2023) 107:12. doi: 10.1007/s10846-022-01781-3

29. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ.* (2025), e081123. doi: 10.1136/bmj-2024-081123

30. Hengartner MP, Jakobsen JC, Sørensen A, Plöderl M. Efficacy of new-generation antidepressants assessed with the Montgomery-Asberg Depression Rating Scale, the gold standard clinician rating scale: A meta-analysis of randomised placebo-controlled trials. *PloS One.* (2020) 15:e0229381. doi: 10.1371/journal.pone.0229381

31. Strauß B, Schumacher J. *Klinische interviews und ratingskalen.* Göttingen: Hogrefe Verlag (2005).

32. Hautzinger M, Keller F, Kühner C. *BDI-II. Beck-depressions-inventar.* Frankfurt: Pearson Assessment (2009).

33. Maher JM, Markey JC, Ebert-May D. The other half of the story: effect size analysis in quantitative research. *CBE Life Sci Educ.* (2013) 12:345–51. doi: 10.1187/cbe.13-04-0082

34. Pszeida M, Häussl A, Zuschnegg J, Macher M, Draxler S, Orgel T, et al. Motivating patients with depression for gender-sensitive cognitive training using a socially assistive robot with bio-signal driven pause management. (2025) 165. doi: 10.54941/ahfe1006090

35. Paletta L, Schuessler S, Reininghaus E, Macher M, Fellner M, Russegger S, et al. Towards gender-sensitive motivation of patients with depression for cognitive training with a socially assistive robot. (2023) 73. doi: 10.54941/ahfe1003969

36. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods.* (2009) 41:1149–60. doi: 10.3758/BRM.41.4.1149

37. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* (2007) 39:175–91. doi: 10.3758/BF03193146

38. Study randomizer. Available online at: https://www.studyrandomizer.com (Accessed July 4, 2025).

39. LimeSurvey GmbH. LimeSurvey: an open source survey tool (2025). Available online at: https://www.limesurvey.org (Accessed July 4, 2025).

40. Aldebaran. Aldebaran - pepper. Marseille, France: Aldebaran (2025). Available online at: https://aldebaran.com/en/pepper/ (Accessed June 6, 2025).

41. Probo Robotics. Technische spezifikationen (2022). Available online at: https://www.probo-robotics.at/pepper-daten-fakten/ (Accessed November 22, 2022).

42. Softbank Robotics. Pepper – The worlds first Personal Robot that Reads Emotions (2021). Available online at: https://www.softbankrobotics.com/emea/de/pepper (Accessed July 4, 2025).

43. Feil-Seifer D, Matarić MJ. (2005). Defining socially assistive robotics, in: *9th Int Conf Rehabil Robot 2005 ICORR*, . pp. 465–8.

44. Paletta L, Fellner M, Pszeida M, Lerch A, Kemp C, Pittino L, et al. (2018). Playful multimodal training for persons with dementia with executive function based decision support, in: *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, Corfu Greece. pp. 237–40. doi: 10.1145/3197768.3197789

45. Paletta L, Pszeida M, Fellner M, Russegger S, Dini A, Draxler S, et al. Towards decision support with assessment of neuropsychological profiles in alzheimer's dementia using playful tablet-based multimodal activation. In: Ayaz H, Asgher U, Paletta L, editors. *Advances in neuroergonomics and cognitive engineering.* Springer International Publishing, Cham (2021). p. 469–77. doi: 10.1007/978-3-030-80285-1_54

46. dZuschnegg J, Schuessler S, Paletta L, Russegger S, Fellner M, Ploder K, et al. Usability evaluation after a 6-month tablet-based dementia training program by people with alzheimer's disease, relatives, and dementia trainers. (2022) 43. doi: 10.54941/ahfe1001844

47. Fluet G, Qiu Q, Gross A, Gorin H, Patel J, Merians A, et al. The influence of scaffolding on intrinsic motivation and autonomous adherence to a game-based, sparsely supervised home rehabilitation program for people with upper extremity hemiparesis due to stroke. A randomized controlled trial. *J NeuroEngineering Rehabil.* (2024) 21:143. doi: 10.1186/s12984-024-01441-7

48. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum-Comput Interact.* (2008) 24:574–94. doi: 10.1080/10447310802205776

49. Brooke J. SUS: A quick and dirty usability scale. *Usability Eval Ind.* (1995) 189.

50. Lewis JR. The system usability scale: past, present, and future. *Int J Human–Computer Interact.* (2018) 34:577–90. doi: 10.1080/10447318.2018.1455307

51. Khan Q, Hickie IB, Loblay V, Ekambareshwar M, Zahed IUM, Naderbagi A, et al. Psychometric evaluation of the System Usability Scale in the context of a childrearing app co-designed for low- and middle-income countries. *Digit Health.* (2025) 11:20552076251335413. doi: 10.1177/20552076251335413

52. Sauro J. *A practical guide to the system usability scale: background, benchmarks & best practices.* Denver, CO: Measuring Usability LLC (2011). p. 162.

53. Kothgasser O, Felnhofer A, Hauk N, Kastenhofer E, Gomm J, Kryspin-Exner I. TUI - technology usage inventory manual (2012). Available online at: https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf (Accessed July 4, 2025).

54. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* (1989) 13:319. doi: 10.2307/249008

55. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. *Manag Sci.* (2000) 46:186–204. doi: 10.1287/mnsc.46.2.186.11926

56. IBM Corp. *IBM SPSS statistics for windows.* Armonk, NY: IBM Corp. (2023).

57. Šabanović S, Chang W-L, Bennett CC, Piatt JA, Hakken D. A robot of my own: participatory design of socially assistive robots for independently living older adults diagnosed with depression. In: Zhou J, Salvendy G, editors. *Human aspects of IT for the aged population. Design for aging.* Springer International Publishing, Cham (2015). p. 104–14. doi: 10.1007/978-3-319-20892-3_11

58. Göransson O, Pettersson K, Larsson PA, Lennernäs B. Personals attitudes towards robot assisted health care – a pilot study in 111 respondents. *Stud Health Technol Inform.* (2008) 137:56–60.

59. Laban G, Kappas A, Morrison V, Cross ES. Building long-term human–robot relationships: examining disclosure, perception and well-being across time. *Int J Soc Robot.* (2024) 16:1–27. doi: 10.1007/s12369-023-01076-z

60. Rosén J, Lindblom J, Lamb M, Billing E. Previous experience matters: an in-person investigation of expectations in human–robot interaction. *Int J Soc Robot.* (2024) 16:447–60. doi: 10.1007/s12369-024-01107-3

61. Mlakar I, Smrke U, Flis V, Kobilica N, Horvat S, Ilijevec B, et al. Using structural equation modeling to explore patients' and healthcare professionals' Expectations and attitudes towards socially assistive humanoid robots in nursing and care routine. *Int J Soc Robot.* (2024) 16:105–24. doi: 10.1007/s12369-023-01039-4

62. Piatt J, Nagata S, Šabanović S, Cheng W-L, Bennett C, Lee HR, et al. Companionship with a robot? Therapists' perspectives on socially assistive robots as therapeutic interventions in community mental health for older adults. *Am J Recreat Ther.* (2017) 15:29. doi: 10.5055/ajrt.2016.0117

63. Becchimanzi C, Iacono E, Brischetto A. Acceptability of assistive robotics by older adults results from a human-centred qualitative study. (2022) 37. doi: 10.54941/ahfe1001637

64. Maranesi E, Amabili G, Cucchieri G, Bolognini S, Margaritini A, Bevilacqua R. Understanding the acceptance of ioT and social assistive robotics for the healthcare sector: A review of the current user-centred applications for the older users. In: Scataglini S, Imbesi S, Marques G, editors. *Internet of things for human-centered design. Studies in computational intelligence.* Springer Nature Singapore, Singapore (2022). p. 331–51. doi: 10.1007/978-981-16-8488-3_16

65. Motahar T, Farden MDF, Sarkar DP, Islam MDA, Cabrera ME, Çakmak M. (2019). SHERA: A low-cost assistive robot for older adults in the developing world, in: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, New Delhi, India. pp. 1–8. doi: 10.1109/RO-MAN46459.2019.8956402

66. Hammer S, Kirchner K, André E, Lugrin B. (2017). Touch or talk?: comparing social robots and tablet PCs for an elderly assistant recommender system, in: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna Austria. pp. 129–30.

67. Mahmoudi Asl A, Toribio-Guzmán JM, Castro-González Á, Malfaz M, Salichs MÁ, van der Roest H, et al. Evaluating the user experience and usability of the MINI robot for elderly adults with mild dementia and mild cognitive impairment: insights and recommendations. (2024) 13:2081–2102. doi: 10.20944/preprints202408.1465.v1

68. Forgas-Coll S, Huertas-Garcia R, Andriella A, Alenyà G. Gendered human–robot interactions in services. *Int J Soc Robot.* (2023) 15:1791–807. doi: 10.1007/s12369-023-01035-8

69. Andriella A, Siqueira H, Fu D, Magg S, Barros P, Wermter S, et al. Do I have a personality? Endowing care robots with context-dependent personality traits. *Int J Soc Robot.* (2021) 13:2081–102. doi: 10.1007/s12369-020-00690-5

70. Ahmad R, Siemon D, Gnewuch U, Robra-Bissantz S. Designing personality-adaptive conversational agents for mental health care. *Inf Syst Front.* (2022) 24:923–43. doi: 10.1007/s10796-022-10254-9

71. Chen F-Q, Leng Y-F, Ge J-F, Wang D-W, Li C, Chen B, et al. Effectiveness of virtual reality in nursing education: meta-analysis. *J Med Internet Res.* (2020) 22:e18290. doi: 10.2196/18290

72. Chen S, Jones C, Moyle W. Social robots for depression in older adults: A systematic review. *J Nurs Scholarsh.* (2018) 50:612–22. doi: 10.1111/jnu.12423

73. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open.* (2018) 8:e018815. doi: 10.1136/bmjopen-2017-018815

74. Eyssel F, Hegel F. (S)he's got the look: gender stereotyping of robots[1]. *J Appl Soc Psychol.* (2012) 42:2213–30. doi: 10.1111/j.1559-1816.2012.00937.x

75. Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: Applications and recommendations for expanded use. *Clin Psychol Rev.* (2015) 35:35–46. doi: 10.1016/j.cpr.2014.07.001

76. Tapus A, Tapus C, Mataric MJ. (2009). The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia, in: *2009 IEEE International Conference on Rehabilitation Robotics*, Kyoto, Japan. pp. 924–9. doi: 10.1109/ICORR.2009.5209501

77. Sharkey A, Sharkey N. We need to talk about deception in social robotics! *Ethics Inf Technol.* (2021) 23:309–16. doi: 10.1007/s10676-020-09573-9