Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learned lessons on Socially Assistive Robot design in Elderly Care: proposal and insights of a Collaborative Design Framework

Antonio Jerez Cordoncillo
Universidad de Málaga
Ana Iglesias
Universidad Carlos III de Madrid
Alberto Tudela Roldán
Universidad de Málaga
Juan Pedro Bandera Rubio
jpbandera@uma.es

Universidad de Málaga

Research Article

Keywords: Socially Assistive Robots, Co-Design, Responsible Design, Human-Robot Interaction, Elderly Care

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10902322/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Learned lessons on Socially Assistive Robot design in Elderly Care: proposal and insights of a Collaborative Design Framework

Antonio Jerez Cordoncillo[1], Ana Iglesias[2], Alberto Tudela Roldán[1], Juan Pedro Bandera Rubio[1*†]

[1*]Department of Electronic Technology, Universidad de Málaga, ETSI de Telecomunicación, Campus de Teatinos, s/n, Málaga, 29071, Andalucía, Spain.

[2]Instituto Agustín Millares & Departamento de Informática, Universidad Carlos III de Madrid, Avda. Universidad, 30, Leganés, 28911, Madrid, Spain.

*Corresponding author(s). E-mail(s): jpbandera@uma.es;
Contributing authors: antoniojcor@uma.es; aiglesia@inf.uc3m.es; ajtudela@uma.es;
†These authors contributed equally to this work.

Abstract

Purpose: This paper proposes a co-design framework and evaluates its practical application during a project involving robot deployment in a retirement home.
Methods: The study details the hybrid methodology used during a 19-week Socially Assistive Robot deployment involving 17 older adults and 3 professionals in a Spanish retirement home. Robot performance was evaluated via the AUSUS framework, prioritizing techniques designed to minimize participant fatigue.
Results: While participants reported high perceived utility, the deployment revealed two distinct obstacles. First, low technical efficacy was primarily driven by inherent unpredictability and daily incidents of a real-world environment, which negatively cascaded into the overall user experience. Second, the study exposed the inherent limitations of traditional User-Centered Design practices, particularly the challenge of defining and sustaining active participant engagement within complex, shared care ecosystems.
Conclusion: Driven by the lessons learned from the CAMPERO project, the paper underscores the necessity for contextual pre-assessments, active participant empowerment, and pragmatic evaluation scoping. Ultimately, these insights

provide a critical foundation for future methodological advancements, ensuring SAR deployments are genuinely collaborative, needs-driven, methodologically responsible and integrated without disrupting the care ecosystem.

Keywords: Socially Assistive Robots, Co-Design, Responsible Design, Human-Robot Interaction, Elderly Care.

1 Introduction

The rapid demographic shift towards an aging global population has placed unprecedented strain on healthcare and socially assistance systems. According to some projections, by 2030, one in six people in the world will be aged 60 years or over [1]. In this context, socially assistive robots (SARs) [2] emerge as a support technology to assist older adults, alleviate the workload of healthcare professionals, and improve overall quality of care. However, despite significant advances in robotic hardware and artificial intelligence, successful, long-term integration of SARs into real-world care ecosystems remains a complex challenge [3]. Often, technological solutions are driven by hardware capabilities rather than the actual needs of the end-users [4], leading to low adoption rates, technology abandonment once the initial "novelty effect" wears off, and a failure to respect the holistic dynamics of the care environment.

To address these challenges, the design and deployment of SARs must evolve from strictly technology-centric paradigms to methodologies that integrate the users and their surrounding ecosystem at every stage. While traditional User-Centered Design (UCD)[5, 6] provides valuable insights, it frequently relegates older adults to a passive, reactive role[7]. Consequently, there is a growing necessity for hybrid methodologies that blend UCD with participatory design techniques. The hybrid approach defined by F. Ashley aims to empower vulnerable populations as active co-creators while strictly managing and mitigating "research fatigue," ensuring that the assessment processes are as respectful of the users' time and energy as the technology itself.

The research work described in this paper presents two primary contributions to address current methodological gaps in Human-Robot Interaction (HRI) within geriatric settings. First, it proposes an hybrid co-design framework and details its empirical evaluation during the CAMPERO research project[1]. Through this "in-the-wild" deployment, a Socially Assistive Robot (SAR) was integrated into the dynamic daily routines of a real care ecosystem.

Second, the paper critically analyzes the operational outcomes, environmental constraints, and direct user feedback from this deployment to derive fundamental methodological lessons. Driven by these empirical insights, the discussion conceptually expands the methodology's approach by highlighting the necessity of contextual pre-assessments, active participant empowerment, and pragmatic evaluation scoping, this critical reflection provides a robust and ethically responsible foundation for the future design and integration of robotic solutions into complex care ecosystems.

[1] https://www.uma.es/avispa/cms/menu/proyectos-de-investigacion/proyecto-campero/

2 Related Work

To contextualize the development of SARs in elderly care, a systematic literature review was conducted covering several databases (IEEE Xplore, PubMed, and SpringerLink) between 2015 and 2026. A total of 38 relevant studies were identified, focusing on the intersection of robotic technology, elderly care, and collaborative design methodologies. The analysis of this literature reveals distinct trends regarding application domains, the degree of user involvement, and testing periods. The gathered data can be seen in tables 1 and 2.

2.1 Shared vs. Private Environments

For this literature review, 5 main group themes have been defined: Aging in Place, Institutional Care, (Physical) Rehabilitation, Cognitive Stimulation and Affective Stimulation. The literature demonstrates a strong inclination towards solutions that promote "Aging in Place." The majority of the reviewed studies (65.78%) focus on independent living at home, proposing SARs to support daily living activities, facilitate communication, and assist unpaid caregivers. Examples of said case are [8, 9]. However, a smaller subset of studies focuses on institutional care and shared environments [10–12]. Deploying SARs in shared ecosystems introduces significant complexities, as the design must account for multiple users simultaneously. Existing literature varies in its focus within these environments, alternating the primary interaction perspective between the resident, the healthcare professional, or other actors such as family members [12]. Other niche but growing applications include physical rehabilitation [13–15], cognitive [16] and affective stimulation [17].

2.2 The Participant's Role: From Reactive Informants to Active Co-designers

According to Sanders [18], design research can be broadly divided into an "expert mindset," where users act as reactive informants, and a "participatory mindset," where users are active co-creators. The current state of the art in SAR development is overwhelmingly dominated by the former.

Out of the 38 analyzed studies, 24 utilized traditional User-Centered Design (UCD) approaches, heavily relying on qualitative techniques such as interviews (the most common, appearing in all of the 24 studies, plus in other 3 with hybrid approaches) and focus groups (appearing in 9 of the 38 studies) to gather requirements [19]. UCD is a concept proposed by professor Donald Norman's research laboratory in the University of San Diego during the 80s [5], and further emphasized and refined in the book *Psychology of Everyday Things* [6].

According to Norman and further researchers exploring this concept [20], the role of the designer is to facilitate tasks for the user, ensuring that they are capable of using the product as intended while employing the minimum effort possible. While these methods are valuable during exploratory phases, they often relegate older adults to a reactive role. Furthermore, while researchers frequently demonstrate predefined

Table 1 Gathered data from the studies (Authors A-N)

Authors	Year	Group Theme	Methodologies	Testing duration
Akinrintoyo et al.	2025	Aging in place	User-Centered Design (UCD), Participatory Design (PD)***	N/A
Bulgaro et al.	2026	Aging in place	Participatory Design (PD)	N/A
Ciuffreda et al.	2025	Aging in place	User-Centered Co-design	N/A
D'Onofrio et al.	2018	Aging in place	Agile Co-creation	N/A
Eftring & Frennert	2016	Aging in place	Iterative Design, Participatory Design (PD)	21 days
Fiorini et al.	2020	Rehabilitation	Iterative Co-creation, User-Centered Design (UCD)	N/A
Fraune et al.	2022	Aging in place	Participatory Design (PD)	N/A
Gao & Kerse	2025	Cognitive stimulation	User-Centered Design (UCD), Participatory Design (PD)	7 days to 90 days
Ghosh et al.	2025	Affective stimulation	User-Centered Design (UCD)	56 days
Gul et al.	2024	Aging in place	Co-design (implicit in experimental session)	N/A
Hing Ting et al.	2017	Institutional care	User-Centered Design (UCD), Participatory Design	4 days
Hsu et al.	2024	Aging in Place	Longitudinal Panel (Long-term Co-design)	180 days
Iglesias et al.	2020	Institutional care	Human-Centered Design (HCD), Participatory Design	540 days
Jerez et al.	2024	Institutional care	User-Centered Design (UCD), Participatory Design	150 days
Juel et al.	2020	Institutional care	Participatory Design (PD)	1 day
Kamino et al.	2023	Aging in place	Co-design	N/A
Lee et al.	2017	Aging in place	Participatory Design (PD)	N/A
McGinn et al.	2020	Aging in Place	Design Thinking	2 days
Naseer et al.	2025	Rehabilitation / Cognitive stimulation	User-Centered Design (UCD), Mixed Methods	56 days and 180 days
Nertinger et al.	2023	Rehabilitation	User-Centered Design (UCD)	2 days

robotic capabilities to users before gathering input to ensure users understand technological boundaries before proposing activities [21, 22], this approach inadvertently limits creativity and restricts the identification of novel, user-driven use cases.

While some studies claim to use "participatory design" [16, 23] to enhance the system's performance and ensure a seamless, needs-driven integration into the care ecosystem, a closer analysis reveals that users rarely maintain an active role throughout the entire process. Instead, the most effective current approaches in the literature utilize a hybrid methodology. In these cases, UCD techniques are used during divergent

Table 2 Gathered data from the studies (Authors O-Z)

Authors	Year	Group Theme	Methodologies	Testing duration
Olatunji et al.	2024	Aging in Place	Immersive Participatory Design	28 days
Olatunji et al.	2023	Aging in Place	Participatory Design	N/A
Olivier et al.	2022	Institutional Care	Co-design (Living Lab)	N/A
Ostrowski et al.	2021	Aging in place	Long-term Co-design (Human-Centered Design)	365 days
Ostrowski et al.	2021	Aging in place	Experience-based co-design	365 days
Ostrowski et al.	2024	Rehabilitation / Cognitive stimulation	Long-term Co-design (1 year)	30 to 365 days
Padir et al.	2014	Aging in place	User-Centered Design (UCD)	N/A
Potter et al.	2026	Aging in place	Phenomenological Co-design	N/A
Šabanović et al.	2015	Aging in place	Participatory Design (PD)	N/A
Sajid et al.	2024	Aging in place	Human-Centered Design (HCD)	N/A
Sienkiewicz et al.	2024	Aging in place	Participatory Design (PD)	N/A
Tokuhisa & Morimoto	2025	Institutional care	Service Design (Ethnographic approach)	N/A
Umbrello et al.	2021	Rehabilitation	Value Sensitive Design (VSD), AI4SG-VSD	N/A
Veling & Villing	2024	Aging in place	"Authentic Quotes" Process (Qualitative analysis)	600 days
Villaverde Naveira et al.	2022	Aging in place	User-Centered Design (UCD)	N/A
Wengefeld et al.	2025	Aging in place	User-Centered Design / Participatory Design	11 to 34 days per user
Wilson et al.	2022	Aging in place	Inductive Design / User-Centered Design	2 to 38 days

exploration phases, while active participatory techniques—which encompass classical software engineering methods for user requirements acquisition, such as design workshops involving card sorting [9], storyboarding [24], and prototyping [17], are employed during convergent, creative phases. Approaches like these, alongside "Authentic Citation" methods designed to prevent user stereotyping [19], prove more effective in aligning the technology with actual user expectations.

2.3 Assessment Limitations and the Novelty Effect

A critical challenge identified in the literature is the duration of SAR assessment periods. While some studies favor long-term deployments (e.g., 150 to 540 days) [10, 11], many rely on short-term evaluations [25, 26].

Short-term assessments are particularly vulnerable to the "novelty effect", which is defined as a temporary spike in user motivation and perceived usability driven by the novelty of the technology. Research indicates that this effect typically requires at least eight weeks to dissipate [27], though this timeframe fluctuates based on variables such as intrinsic motivation, intensity of use, and the introduction of new features during iterative design [28]. Consequently, short-term evaluations involving vulnerable populations risk yielding biased data that, while accurate in the moment, do not reflect realistic, long-term technological adoption.

To account for the novelty effect, the 38 identified studies were classified into three distinct cohorts based on their longitudinal scope, utilizing an eight-week threshold as the differentiating criterion. These categories comprise: (i) long-term deployments with an evaluation period greater than 8 weeks (n=9), ii) short-term evaluations lasting lesser than 8 weeks (n=9) and (iii) studies lacking a solid empirical assessment phase (i.e. brief sessions with short interactions) (n=19).

3 Methodology

This section details the design framework employed for the CAMPERO project. The methodology adopted for this research project was hybrid, combining User-Centered and Participatory Design methods. However, most methods fell under the former category. Although the user adopted a reactive and informative role (typical of UCD methods) in most design phases, the design team prioritised equipping them with all available tools and information to empower them throughout the process

The structure of the co-design framework encompasses 4 stages, following the co-creation definition of Smallman and O'Donovan [29]: "a process that aims to improve efficiency, effectiveness and acceptability by involving stakeholders in identifying, developing, implementing and assessing products, services, policies and systems".

As Table 3 shows, these four phases are analogous to other design processes' structure in different research fields. The development of novel methodologies is often perceived as a break with previous paradigms. However, the most robust innovation is usually built upon universal structures. By proposing a specific co-design framework for social assistive robotics, it becomes evident that the underlying architecture (i.e. identifying, developing, implementing, and assessing) does not seek to reinvent the wheel, but rather respects an iterative philosophy widely consolidated in computer science and systems design. For this study, four main system designs from different fields have been taken as a reference to structure this co-design framework, shown in Table 3. These are: Analysis, Design & Development, Implementation and Evaluation (ADDIE), a framework for building programs [30], Software Development Life Cycle (SDLC) [31], Quality Management System (QMS) [32] and Design Thinking [33].

Given that the operational core of this framework revolves around the programming of use cases for socially assistive robots, its structure naturally inherits the

philosophy of the Software Development Life Cycle (SDLC). However, programming a robot designed to interact with humans in caregiving environments is not equivalent to deploying traditional software, as the amount of environmental variables is much higher, thus incrementing the complexity of the design itself. This is where the proposed co-design framework acquires its specialized character.

Therefore, this co-design framework requires transforming the traditional SDLC phases to absorb the complexity of the elderly care ecosystem. For instance, the "Identifying" (Analysis/Requirements) phase must transcend the mere collection of user specifications to integrate imperatives of User Centered and Participatory design.

Similarly, the "Developing" phase is not limited to writing efficient code, but rather to the choreography of adaptive social interactions. Finally, the "Implementing" and "Assessing" stages must be redefined: it is not solely about verifying the absence of execution errors, but about measuring the ecosystemic impact of the robot in its real-world care environment, ensuring a responsible and safe interaction.

In conclusion, this co-design framework acts as an interdisciplinary bridge. It takes the structural and predictable rigor of software engineering and provides it with the human sensitivity, co-design methodologies, and ethical responsibility necessary to face the unique challenges posed by social robotics in the real world.

Additionally, an "in the wild" approach was followed, meaning that, aside from the technical development of the robot, the rest of the activities were performed in a real environment. The identifying, implementation and assessment phases were conducted in the retirement house "Vitalia Home Teatinos", in Málaga, Spain. During the whole design process, both residents and professionals were involved in said phases, shaping the functionality of the SAR according to their needs.

Table 3 Comparative of phases from different design methodologies

Phases	ADDIE [30]	SDLC [31]	QMS [32]	Design Thinking [33]
Identifying	Analysis	Investigation & Analysis	Diagnosis	Empathize & Define
Developing	Design & Development	Design & Development	Planning	Ideate & Prototype
Implementing	Implementation	Implementation	Implementation	Test
Assessing	Evaluation	Management	Evaluation & Reflection	Feedback & Iteration

3.1 Participants

For this research work, a total of 17 residents (Female $N = 12$) between 59 and 95 years of age ($M = 81.25$; $SD = 10.17$) participated in the study. Furthermore, during the identifying and implementing phases of the design, various healthcare professionals were involved (psychologist, physiotherapist and occupational therapist). The recruitment process was performed by the professionals of the retirement home, following the

selection criteria defined by the research team. In this case, the main exclusion criteria was declining cognitive function, which was measured by the psychologist of the retirement house using the Lobo's Mini-mental scale [34]. Although our population is varied age-wise, thus allowing for different score thresholds, a minimum punctuation of 24 was established by the professionals. This allowed the research work to have a varied population physical and sensory-wise, which helped the research work to test the adaptive capability of the SAR. All participants in the research work signed an adequate informed consent.

Furthermore, an *ex ante* evaluation (i.e. before the SAR's deployement) is conducted, which was divided in two parts: a demographic and technology use questionnaire where general information about the participant is recollected, and a second questionnaire which recollected parameters regarding declining sensory functions (sight and hearing) and motor skills (i.e. whether they use wheelchair, a walker or a cane). This questionnaire was also the only one related to the participants which was filled by the healthcare professionals, due to sensitivity reasons, and because caregivers have a more objective view and more details regarding the declining sensory functions and motor skills of the residents.

The descriptive statistics, including the mean (M) and standard deviation (SD) for each item ($N = 17$), are presented in Table 4. Evaluated on a 5-point Likert scale (1 = Any issues, 5 = Severe, for the Parameter's items, and 1 = Never, 5 = Always, for the Previous Technological Experience's items).

Table 4 Baseline descriptive statistics for declining sensory functions and motor skills and technology use ($N = 17$).

Category	**Questionnaire Item**	**Mean (M)**	SD
Parameters	Declining sensory function (Sight)	1.941	0.827
	Declining sensory function (Hearing)	1.765	1.214
	Declining motor skills	2.824	1.042
Previous Technological Experience	How often do you use your mobile phone?	2	0.767
	How often do you use your computer?	1.294	0.749
	How often do you use your tablet?	1.059	0.235

3.2 Identifying

The goal of this phase is to identify the needs of the users in order to define the use cases. Therefore, a study of the ecosystem that surrounds the primary user is conducted.For this scenario, an User-Centered Design approach is used which combines two techniques: focus groups and storytelling. These two techniques allow obtaining a broad picture of the day-to-day life, not only for the primary users, but for the other

actors in the ecosystem. The focus group will answer 5 questions following a round-robin order, which will help the design team to create a picture of the ecosystem. These questions are:

- Tell us what your day-to-day routine is like.
- What are the most challenging moments of your day?
- What are the most enjoyable moments of your day?
- What technologies are used in the retirement house?
- Which functionalities do you think the robot can have to be helpful in your environment?

Following the structured inquiries, an open-ended discussion period is provided for participants to contribute additional insights. This interactive process serves a multi-faceted purpose:(i) breaking the ice, facilitating rapport and initial engagement; (ii) obtaining a general idea of the participant's daily routines; (iii) detecting moments of the day when there is some difficulty that affects the user's quality of life; (iv) detecting moments of the day that are essential for the user (these moments should be favored and never eliminated once the technology is implemented); (v) knowing more about the digital aptitudes within the care environment; and (vi) exploring participants expectations regarding the use of a robotic platform in care environment.

The qualitative data gathered from these focus groups were subsequently synthesized into a structured list of user's requirements, which served as the foundation for the use cases detailed below.

3.3 Development

As mentioned previously, the development phase is the only phase of the design process which was not conducted alongside the end-users. Instead, it was built within the collaboration of the technical team, an anthropologist and a sociologist. Before proceeding with the use case definitions, it is important to highlight that, as the SAR's dynamic is based in face detection to enhance its personalized interaction, pictures and parameters data recollected during the ex ante evaluation were stored in the SAR's database.

Use case 1: Wandering

Basic use case of the SAR. It navigates autonomously within the preestablished zone, avoiding obstacles and detecting people. As soon as it detects someone whose data are in the SAR's database (i.e. those residents participating in the test sessions), it begins an use case, or several of them in a chain, depending on the detected person. Furthermore, during this state, the SAR will display on the screen the default interface showed in Fig. 1. Following user's requests, a slider in the top center was added to easily change voice volume for the robot. Finally, the battery level will always be displayed, even in other interfaces (e.g. in the *Menu Selection* interface shown in Fig. 2).

Use case 2: Menu Selection

This use case starts when the robot encounters an user that is authorized to provide her menu options, but didn't do it yet. Once these initial conditions are met, the robot



Fig. 1 Initial (default) Interface

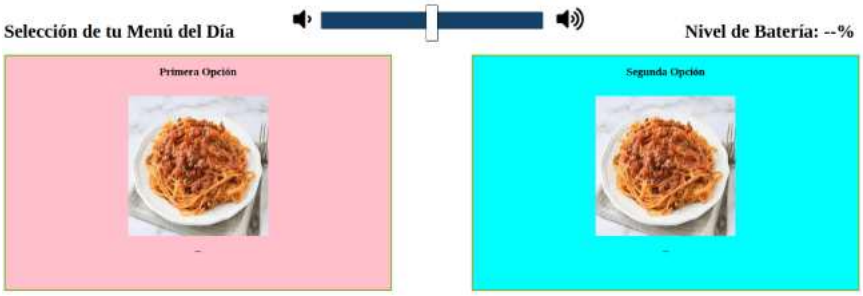


Fig. 2 Menu Selection Interface

will reposition itself within a certain distance of the user. This distance depends on the particular user profile to facilitate the interaction, so the robot can adapt for users that need a walker, or wheelchair, or that are seated, or simply users that want the robot to stay closer or farther away. The SAR will announce the beginning of the use case calling the user by her name. Then, the daily menu options for the next week will be displayed, along with illustrations that represent each of the options (Fig. 2). Collected results are sent to the kitchen staff via email.

Use case 3: Cognitive stimulation

This use case can only be executed during the scheduled cognitive stimulation activity time. During this period, once the SAR recognizes an older adult with prior professional authorization to perform the cognitive stimulation activity, it will announce the beginning of the use case calling the user by their name. Then, a sequence of questions will be displayed in the robot's screen. Each question has 4 possible answers as shown

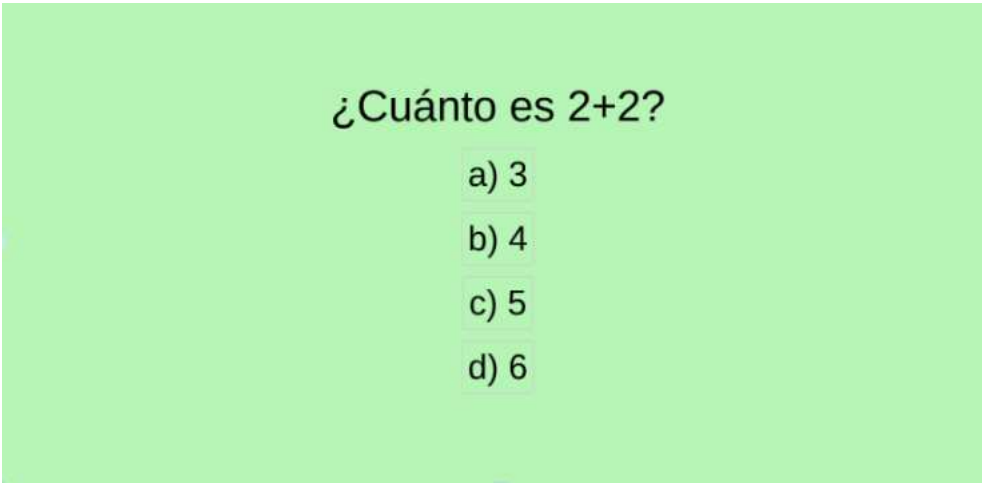


Fig. 3 Cognitive stimulation Interface

in figure 3. Once the user chooses one of the answers, the SAR will inform the user whether they were right or wrong via the screen. After answering all the questions, the results are sent to the psychologist of the institution via email.

Use case 4: Reminder

If the SAR detects a user who needs reminding, it will start this use case, always calling them by their name. This reminder adapts to any particular circumstance (visit of relatives, appointments, individual therapies, etc.), and it is also configured depending on the user profile to adapt to the personal preferences of the user.

Use case 5: Musical Therapy

Once the predetermined session time comes, the SAR will navigate to the specific position within the room where the therapy is hold. There it will play a specific audio (generally music chosen by the users or the therapist in charge of the session). At the same time, the SAR will perform small movements, simulating a dance.

Use case 6: Bring me water!

This use case is the only one performed in a reactive manner. It starts when the user presses the "Request" button included in the upper left corner of the default interface (Fig. 1). Once the button is tapped, the SAR must recognize the user.

If the user is successfully identified, the SAR will display 4 options, as shown in Figure 4. Depending on the user's need, they would be able to choose one of these. Then, the SAR will navigate to the room entrance, playing an audio which describes the need of the user.

If the user is not successfully recognized, the SAR informs it cannot perform the requested action, and resumes use case 1: wandering.

Subsequently, these formal Use Cases are systematically decomposed into a predefined set of software functionalities and interaction specifications for the SAR.

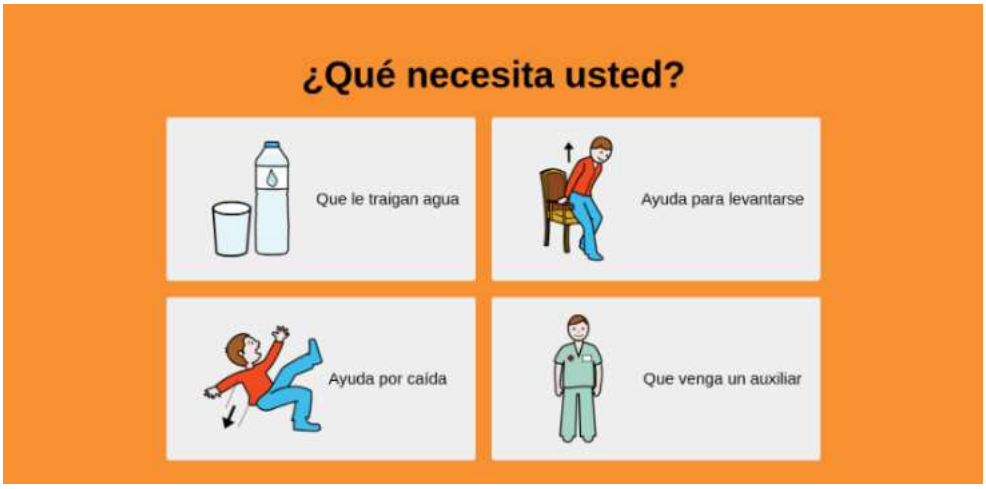


Fig. 4 "Bring me water!" Interface

From the use cases defined for this research work, the following five functionalities were set as mandatory for the SAR to have:

- Autonomous navigation.
- Screen interaction.
- Gather the data received through the screen interaction.
- Provision of audio answers for the user.
- Provision of screen answers for the user.

These high-level functionalities are then further broken down into discrete functional and non-functional requirements. Each requirement is categorized as mandatory or optional and is precisely defined to ensure it can be objectively evaluated against specific metrics within an experimental plan. Specifications are built over these functionalities as lists of functional and non-functional requisites, mandatory or optional, precisely defined and able to be objectively tested following certain metrics and experimental plan. For instance, the functionality 'navigate autonomously' will decompose into a set of specifications such as: (Functional/Mandatory) The robot locates itself in a map of the retirement home, using only onboard sensors; (Functional/Mandatory) The robot can set a path to reach a certain pose in the map, defined as (x,y,theta) coordinates; (Functional/Mandatory) The robot can dynamically modify its path to avoid collision with unpredicted obstacles; (Functional/Optional) The robot will interact with people when they are blocking its path, kindly asking them to let it pass; etc. During the development of the project, two main robotic platform were available for the software development of the use cases: CLARA and GoBe robot (both used in previous studies [11]). Ultimately, the CLARA robot was chosen due to fulfilling the aforementioned functionalities. Once the chosen robotic infrastructure is programmed with the defined use cases, the Development phase is finished, therefore leading to the next one: Implementation.

3.4 Implementation

The SAR was deployed in the retirement house between June 14th and November 22nd, 2024. However, during the month of August, no tests or evaluations were performed due to the holiday period, resulting in a total assessment time of 19 weeks.

Although this phase carries a lower methodological load than the preceding stages, it is of vital importance for the success of the overall design. A flawed implementation can precipitate technology rejection by the users and heavily bias the outcomes of the Assessment phase. Therefore, this design framework establishes three key factors to ensure a successful and bias-free implementation.

Key Factors of the Implementation

- Deployment period: It must be sufficiently long to allow the "novelty effect" to dissipate before the final assessment phase, ensuring that evaluation data reflects realistic, long-term adoption rather than temporary intrigue.
- "In the wild" testing: For a realistic implementation, the SAR must be tested in a real-world environment with actual end-users. While logistically complex, only testing in these environments exposes the contextual nuances and unpredicted variables that remain hidden in conventional laboratories or virtual simulations.
- Training and onboarding (Co-implementation): The design team must carefully manage the technological learning curve. Insufficient training can induce user frustration, skewing the final results. To mitigate the fear and uncertainty often experienced by older adults when facing novel and abstract concepts [35], the co-design framework advocates for "co-implementation." This approach involves introducing the technology through familiar actors within the care ecosystem (e.g., healthcare professionals, relatives, or peers of a similar demographic [36]). This collaborative strategy brings the technology closer to the target audience in a more organic, natural, and calming manner.

Session Zero and User Testing Execution

Translating the co-implementation strategy into practice for this research work, a healthcare professional actively assisted during the initial sessions to facilitate the residents' acceptance and learning process. As part of this onboarding, a "Session Zero" workshop was conducted to formally introduce the technology to the users. The structure of this workshop was as follows:

- Presentation of the moderators and the robot.
- Brief description of the intended use cases.
- An open discussion session with the participants regarding the use cases.

Following this initial workshop, testing sessions were conducted almost daily in the retirement house until the conclusion of the pilot project. During these sessions, available participants gathered in one of the common rooms while the SAR initiated its baseline state: Wandering (Use Case 1). From there, relying on contextual triggers and the specific identities of the present participants, the SAR dynamically activated the subsequent use cases. During the 19-week period, a total of 70 sessions were performed.

3.5 Assessment

The assessment of the research work was planned to be conformed in two parts: the aforementioned ex ante evaluation (i.e. before the SAR's deployment) which recollects demographic data from the participants (age and gender), the grade of declining sensory functions and motor skills, and previous technological experience; and a ex dure and ex post (i.e. during and after the deployment, respectively) evaluation which recollects data regarding several aspects of Human-Robot Interaction through the evaluation framework of AUSUS [37].

This evaluation framework, already employed by the authors in previous works [11, 37, 38], was used to evaluate several aspects of Human-Robot Interaction. However, the methods used for this evaluation changed from previous iterations. The methods used to assess each of the chosen indicators are shown in Tables 5 through 8. Due to the long-term characteristic of this pilot and the nature of the participants, the evaluation methodology was designed with a main goal: to reduce the "research fatigue" as much as possible. This means designing UCD activities, such as interviews or questionnaires, as short and simple as possible, but always gathering as much data as possible. Other research methods were also incorporated, such as focus groups and expert evaluations. Furthermore, in the case of the focus groups, a more participative approach is chosen, letting participants assume a more active role (more details in the Embodiment subsection). It is important to highlight that not all indicators from the AUSUS evaluation framework were chosen, and only those which were key for the research were considered. As most of them are evaluated through questionnaires, the chosen indicators were carefully selected.

Although the initial evaluation plan included two rounds of questionnaires, one around the middle of the pilot and another at the end of it, the latter could not be completed, as most participants at that point of the pilot lost interest in the study or were not able to go through the structured interview. However, one successful method that provided useful data were the brief unstructured interviews where the researcher, every once in a while, asked the participants (whenever they were available) any opinion about the robot. That method became the main feedback provider during the last phase of the experiment.

Table 5 Assessment methodology: Accessibility

Indicator	Method
Perception	Questionnaire
Operation	Questionnaire
Understanding	Questionnaire
Robustness	Expert Evaluation

Table 6 Assessment methodology: Usability

Indicator	Method
Effectiveness	Expert Evaluation
Efficiency	Expert Evaluation
Learnability	Expert Evaluation
Flexibility	Expert Evaluation
Robustness	Expert Evaluation
Utility	Questionnaire

Table 7 Assessment methodology: Social Acceptance

Indicator	Method
Effort Expectancy	Questionnaire
Attitude toward Using Technology	Questionnaire
Self Efficacy	Questionnaire
Attachment	Questionnaire
Reciprocity	Questionnaire

Table 8 Assessment methodology: User Experience

Indicator	Method
Embodiment	Focus group

4 Results

The results of the evaluation of this research work are organized in two main sections: i) the AUSUS evaluation; and (ii) a section called "Other aspects", in which additional qualitative data are discussed.

4.1 AUSUS Evaluation

As with the ex ante evaluation, the passed questionnaires follow a 5-point Likert Scale. The inquiries conforming said questionnaire are shown in Table 9.

As noted in the Assessment section, participant attrition during the final weeks of deployment limited the formal AUSUS evaluation to a single round of administered questionnaires, rather than the initially planned two. Nevertheless, supplementing this formal assessment with the aforementioned short, unstructured inquiries conducted post-session yielded a substantial amount of qualitative data. This highly adaptive data collection approach complements the rest of feedback collection procedures to finally provide sufficient empirical evidence to derive robust conclusions, evaluate the impact of the SAR, and refine the proposed co-design framework.

Table 9 Data from the questionnaires

Indicator	**Question**	**Mean (M)**	SD
Perception	I could easily hear what the robot said	4.500	0.707
Perception	I could see what the robot showed on the screen	4.429	0.904
Operation	I could interact with the robot through the screen	2,833	1,067
Understanding	I understood what the robot communicated	3.875	1.452
Utility	I think the robot is useful in the retirement home	3.875	1.364
Effort Expectancy	I find the robot easy to use	3.875	1.269
Attitude towards technology	I have a positive opinion regarding the use of robots in the retirement home	4.125	0.927
Self-Efficacy	I feel capable of performing my daily tasks	4.125	0.927
Attachment	I would like the robot to stay at the retirement home	3.625	1.654
Reciprocity	I believe the robot responds appropriately	3.625	1.576
Human Oriented	Talking to the robot feels similar to talking to a human	2.875	1.615

Efficacy

During the 70 sessions performed during the 19-week period, 165 use cases were registered (i.e. complete user interactions defined previously), although only 71 of them were successfully finished.

In order to measure the efficacy of the SAR, the number of successful use cases was divided by the number of total use cases. In this research work, the efficacy of the SAR was 43,03%.

Most of the failed use cases were caused by errors in the face recognition system, that either mismatched the user or did not recognize them at all, due to direct sunlight entering the common room where it was working. All sessions were performed during the morning, and the curtains were left open because participants preferred them that way. Reflections produced noisy laser scans. Furthermore, the SAR also had some navigation problems due to the furniture's position in the common room being constantly changed. The combination of these two factors made the performance of the SAR be lower than expected.

Another factor preventing the success of the use cases was the tactile screen. Even when the screen was specifically selected for its good sensitivity, and during lab tests it worked fine when operated by people used to these devices, in the experiments in the retirement home the tactile screen revealed as a very difficult to operate interface for the residents. Through user observation, the two main reasons behind this difficulty are i) limited digital skills and ii) age-related tremors.

Ultimately, these environmentally driven technical failures acted as a severe bottleneck for the human-robot interaction process. Consequently, the low efficacy rate (43.03%) must be treated as a primary confounding variable when analyzing the subsequent subjective HRI indicators, as the environmental limitations directly cascaded into user frustration and lower perceived usability.

Efficiency

Efficiency is a critical dimension of usability, measuring the resources and time expended by users to achieve their goals during interaction. Originally, this indicator was intended to be comprehensively evaluated by recording the average duration of successful instances for each use case. However, as detailed in the previous section, the severe environmental bottlenecks and the correspondingly low efficacy rate (43.03%) drastically reduced the pool of successfully completed interactions.

Consequently, while efficiency remains a vital metric within the proposed evaluation framework, the empirical data collected during this deployment was not sufficiently robust to draw clear, generalizable conclusions. The limited timing data extracted only reflects an idealized subset of interactions that managed to bypass environmental disruptions, thus failing to represent the true longitudinal experience of the users. Therefore, a rigorous evaluation of the SAR's efficiency could not be properly concluded, highlighting a crucial methodological reality: assessing interaction efficiency in "in-the-wild" settings is entirely contingent upon first securing a reliable and robust baseline of technical efficacy.

Perception

As Table 9 depicts, most of the participants were able to hear the robot ($M = 4.5$). Some users encountered some difficulties due to having declining hearing functions or due to the audio quality ("There is noise in the [robot's] audio") (Participant 12). In another case, one of the participants stated that "I don't like it [the sound of the robot's voice]" (Participant 1).

On the other hand, regarding the sight perception, similar results were collected, where the few participants who were not able to see what the robot was displaying on the screen were the ones with declining sight functions, or, in one of the cases, with watery eyes caused by dry eye symptom.

Finally, participants were asked what could be done so they could retain the information easier. The most common response were "higher volume [for the audio]" and more use of "drawings" on the screen.

Operation

This indicator has the lowest score among all the items assessed in this research work (M=2.833). Five participants stated that the main issue they had while interacting with the robot was the tactile screen, which was not working as they expected and was difficult to control. For instance, some of them will tap the screen for too long or for too short so the tactile input was not recognized. One participant complaint about the speed of the robot's movement, which did not allow her to properly interact with it.

Understanding

For this item, a clear disparity was found. While most of the participants understood what the robot was saying/asking to the user the whole time, two of them did not, and both for the same reason: they were not "paying attention". This result, as isolated as it is, brings several thoughts into this research. First, as "flashy" or "eye-catching" a robot in a retirement home can be, these two participants treat it like something ordinary, which can be easily blend into the environment. This is not the first time a robot has been navigating through this specific environment, as it has been used in previous research works [10, 11], so some residents are starting to get used to see robots wandering around, thus minimizing the novelty effect to a degree. And then this situation can make these residents not to pay attention to the SAR, even when it is trying to address them in HRI processes.

Utility

Utility is related to how a device can be used to reach a certain goal or to perform certain task [39]. The general consensus regarding the utility of robots in retirement homes is positive as seen in Table 9 ($M = 3.875$, $SD = 1.364$). Residents are aware of the needs in such care environments, and, as it will be discussed in the next section, some of them see SARs as useful and valuable tools, not only for them, but also for the professionals. Qualitative feedback further supported this metric, indicating that residents are acutely aware of the logistical needs within their care environment. For instance, Participant 3 explicitly praised the robot for its "utility and simplicity", highlighting the importance of accessible and straightforward design. Furthermore, Participant 4 reinforced this pragmatic perspective by asserting that "its (the SAR's) goal is to be useful". These remarks demonstrate that older adults often evaluate technology through a highly functional lens, valuing SARs as practical tools intended to assist both the residents and the healthcare professionals.

Attitude towards the use of technology

According to the UTAUT model, the attitude toward using technology is defined as "an individual's overall affective reaction to using a system" [40]. Regarding this indicator, mixed answers were found. Most of the participants who responded to the questionnaire stated positive attitudes such as "It (the robot) is necessary" (Participant 3), "I like him [the robot]" (Participant 5), "It is entertaining" (Participant 6) or "It is useful, specially the "Bring water" function" (Participant 14). However, there is a minority who are both reluctant to new technologies, stating that "it is going to steal the job of people" (Participant 1), or who believe that "older people are not interested in learning to use new technologies" (Participant 4). Nevertheless, the participant who stated the latter also values the support that these technologies provide to the healthcare professionals. Aspects

Effort Expectancy

Effort expectancy is defined as to which extent the user anticipates a system to be easy to use [39]. Although most of the users were not always successful during the interactions, they perceived the SAR easy to use. However, it is important to highlight

that some participants experienced frustration when the SAR did not respond properly to the inputs they gave (for instance, when tapping the screen in various use cases). It is essential to contextualize that the frustration experienced by the participants was not caused by cognitive complexity (i.e. the interfaces were intuitive and designed with bright colors), but rather by the hardware issues that did not allowed the users to use the SAR properly, thus disrupting the intended ease of use.

Self efficacy

Self efficacy is defined as the person's perception or belief of their ability to reach a goal [39]. For this study, this indicator has been interpreted as the person's perception of their ability fulfilling their day-by-day tasks with a SAR in the same environment. According to the score in Table 9, the participants felt that they were able to fulfill their daily activities with the SAR performing its use cases around the retirement home. This result not only reflects that the SAR functionality is aligned with the necessities of the users within this environment, but also shows that it does not hinder or obstruct any of their daily activities, which is one of the goals that the initial focus group had.

However, this indicator can be also interpreted in this context as the person's belief of their ability to fulfill the tasks asked to do while interacting with the SAR. Although this interpretation was not the one which was focused in this study, later during the Learnability section, a special case will be discussed regarding this indicator.

Attachment

With the exception of one of the participants who did not wanted the robot to be in the retirement home, the feelings of the rest were divided into people that were indifferent and people who wanted the SAR to stay. In the case of the latter group, statements such as "it is a distraction and is something original. But I want it to be improved so it is able to see me" (Participant 16) reflect how this kind of shared environments fall into a state of monotony and when there is a new element (specially when is able to behave differently every day) it is highly valued, not only for its utility, but also for the innate entertainment it offers. This is further proved with Participant 13's opinion, in which they describe the robot as "funny"

Other quote that caught the attention of the researchers was "I like when the robot calls me by my name" (Participant 3). Furthermore, this participant also stated that "the robot is growing on me". Regardless of the few successful use cases, whenever the SAR managed to detect the participant correctly, thus being able to address them by their name, the participant could not help feeling joy and satisfaction. Not only due to the success of the use case, but also because something, albeit knowing that the SAR is not a real person, called her by her name. This situation rises the question whether designing the robot to be able to call the users by their name could be considered ethically correct, as this gesture can develop emotional attachment with the SAR. Further research is needed on whether this result is beneficial or detrimental for the user.

Reciprocity

This indicator was not trivial to assess due to the low success rate of the use cases. Although having a decently high score ($M = 3.625$), this score only reflects the successful cases. So it is logical to think that the score reflects instances where the robot reacts correctly and politely. However, due to the lack of effectiveness, frustration among the participants has been reported during the third month of the pilot. Furthermore, for the open question "Which other use cases that the robot could perform?", one of the participants did talked about a more realistic interaction, through a "more human treatment (maintaining small conversations)" (Participant 12).

Human-Oriented Perception

This indicator was evaluated by measuring the capability of the robot to track human faces and adapt to the specific needs of the user in front of it. As this characteristic was part of the main goals while designing the SAR, the Human-Oriented Perception is really related to the Efficacy.

Learnability

As commented during the Effort Expectancy section, the participants perceived the SAR easy to use. Thus, most of the participants (specially those who attended the testing sessions more frequently) learned to use the SAR easily. However, one of the participants (Participant 5) struggled using the tactile screen from the beginning, experimenting a certain degree of frustration. It was not until month 3 when this participant improved their skills with this kind of screens, which derived not only in a better performance during the SAR interactions, but also a better sense of self-efficacy.

Embodiment

Unlike the rest of the indicators, the embodiment of the robot was assessed via focus group. In this case, a more participatory approach was adopted, similar to a workshop, where a group of photos of different robots (both extracted from fictional and real world) were shown to the participants, who selected their three favorites. The goal behind this new approach was to detect which embodiment aspects of a SAR (color, shape, size, etc.) were fundamental for its acceptance. Initially, adding the CLARA robot to the options was considered, but in order to avoid a biased answer (as by then some of them were already attached to this robot), it was ultimately ruled out.

Seven women and one man participating in the pilot took part in this workshop. Also, one of the healthcare professionals participated as well, but as support for the design teams, in case they need some assistance with the residents. The results were mostly one-sided: while most women chose robots with clean and smooth designs (such as Temi [2] or Nao [3]), the man chose a less appealing option (such as C3PO), comparing it to a "soldier". In second and third positions other robots were chosen due to its utility, despise of not being that "friendly" (such as floor-cleaning robots).

[2] https://www.robotemi.com
[3] https://maxtronics.com/en/nao6/

As a significant secondary finding derived from evaluating this indicator, empirical observations demonstrated that focus groups employing a participatory approach successfully yielded extensive qualitative data without inducing research fatigue. When appropriately structured, the participative approach was perceived by older adults not as an assessment burden, but as a ludic and socially engaging activity. Furthermore, by anchoring the session around open-ended inquiries, participants were granted the agency to guide the peer-to-peer discussion. This conversational autonomy not only minimized cognitive strain but actively fostered intrinsic motivation. Although certain participants initially required encouragement to engage in the group dynamic, the collaborative facilitation provided by both the healthcare professionals and the design team effectively foster their participation. This joint support ensured an inclusive environment where even the most hesitant individuals were able to confidently articulate their perspectives, transforming a traditionally demanding research task into an empowering and highly accepted collaborative experience.

Beyond the structured focus group, continuous qualitative feedback provided valuable insights regarding the SAR's physical and social embodiment. The baseline design was positively received on an affective level, with Participants 5 and 13 explicitly describing the robot as "friendly". However, this initial acceptance was accompanied by a clear desire for enhanced anthropomorphic traits to facilitate a more natural human-robot interaction. Specifically, Participant 13 expressed the need for a "more human appearance", while Participant 3 highlighted the importance of the robot exhibiting more "facial moves". These remarks indicate that while older adults may pragmatically accept clean and functional designs, integrating subtle, expressive human-like features remains a critical factor in designing SARs.

4.2 Discussion on open-ended answers

Lastly, one of the participants commented: "I am aware of the situation [the SAR not working properly] and I appreciate the effort you guys are doing" (Participant 3), during one of the sessions when the technical team has to come to the retirement house in order to fix the SAR. Although seemingly unrelated to the study, this comment provides a huge insight, not from an user experience perspective, but from a design one. This research study, as many other based in UCD, are characterized by their iterative nature: as constant feedback from the participants is informing the technical team about the SAR, the latter need to be active when any changes or fixes are needed. This necessity is even more pronounced in "in the wild" studies, as technology is more prone to stop working properly or even break. This is the reason why the comment showed in this section has so much importance: not only shows the participant's awareness about the work of the technical team, thus helps to leverage the high expectations from the SAR, but also shows a certain degree of resilience within a research context, which matters specially in long term pilot such as this research study.

5 Lessons learned

The deployment of the SAR during the CAMPERO project served as a rigorous empirical test for the proposed co-design framework. The operational outcomes,

environmental constraints, and continuous user feedback highlighted the inherent complexities of "in-the-wild" deployments, exposing critical vulnerabilities in traditional research techniques, such as UCD methods. This section discusses the fundamental conceptual lessons derived from this deployment, underscoring the paradigm shifts required for future robotics research in care ecosystems.

Lesson 1: Importance of the contextual environment

A primary lesson from the CAMPERO deployment is that the success of a SAR is dictated more by the physical and logistical realities of the care ecosystem than by the robot's inherent software capabilities. The majority of performance failures observed, such as natural light blinding sensors or dynamic furniture repositioning hindering navigation, stemmed from environmental unpredictability. This reality highlights the absolute necessity of conducting contextual pre-assessments before any technological intervention. Future methodologies must prioritize mapping the environmental constraints and the daily workflow of healthcare professionals, recognizing that forcing a vulnerable ecosystem to accommodate technological assumptions will inevitably lead to low efficacy and user frustration.

Lesson 2: Necessity of active engagement

Addressing participant attrition is arguably the most significant logistical challenge in longitudinal studies. Initial observations during the CAMPERO project revealed that when older adults are positioned merely as reactive informants (a common limitation of traditional UCD) longitudinal engagement wanes significantly. The study demonstrated that research fatigue is not exclusively a product of the time invested, but fundamentally relates to the quality of the participants' involvement. The necessity lies in transitioning older adults from passive subjects to active co-designers. Sustaining long-term engagement requires methodologies that prioritize active empowerment and conversational autonomy, effectively offsetting the cognitive load typically associated with traditional evaluation methods. As seen during the CAMPERO project, methods such as participative workshops, as cognitively loaded as they might initially appear, demonstrated that granting agency and providing a dedicated space for open feedback did not exacerbate research fatigue.

Lesson 3: Pragmatic scoping

Finally, the deployment revealed a critical dissonance between theoretical methodological design and practical execution. For the CAMPERO project, the evaluation protocol was intentionally designed *a priori* to be as lightweight as possible, aiming to minimize participant burden without sacrificing academic rigor. However, despite these preventative efforts, the methodology still proved overly extensive when applied in the wild.

Consequently, evaluation methodologies must prioritize longitudinal viability over sheer quantitative volume. When selecting indicators to be assessed via questionnaires or structured interviews, research teams must prioritize analytical focus over quantitative volume, refraining from evaluating an excessive number of metrics. Instead, strict pragmatic scoping must be applied, deliberately limiting the assessment to a

concise set of core indicators. While this deliberate reduction inherently narrows the overall scope and ambition of the study, it is a necessary compromise, as it ensures that evaluating the user experience remains genuinely feasible and sustainable across long-term, longitudinal deployments without inducing research fatigue.

6 Conclusion and Future Work

The successful integration of Socially Assistive Robots (SARs) into real-world elderly-care environments requires methodologies that transcend traditional laboratory-centric engineering. This paper presented the deployment of a SAR within the CAMPERO project, serving as a empirical evaluation of the co-design framework proposed in this research study.

By analyzing the operational outcomes, environmental constraints, and continuous user feedback, this study uncovered critical vulnerabilities in methodological approaches and highlighted fundamental paradigm shifts necessary for "in-the-wild" robotics. Firstly, the deployment demonstrated that a robot's success is deeply tethered to the physical and logistical realities of its ecosystem, necessitating contextual pre-assessments prior to any technological intervention. Secondly, the findings challenge traditional notions of research fatigue; to sustain longitudinal engagement and avoid participant attrition, older adults must be empowered as active co-creators rather than relegated to passive informants. Lastly, the study underscores the absolute necessity of pragmatic scoping. To ensure longitudinal viability, research teams must limit the assessment metrics to a concise, comprehensible core to avoid overwhelming vulnerable users and introducing response bias.

Ultimately, these empirical lessons provide a conceptual foundation for the evolution of Human-Robot Interaction (HRI) methodologies in gerontology. Furthermore, the learned lessons in this research work are intrinsically linked to the context in which the co-design framework has been employed, thus it may be not be applicable to other environments. Future work will focus on synthesizing these qualitative reflections into a structurally refined and formalized iteration of this co-design framework. This subsequent methodological proposal will then undergo empirical validation in new residential care ecosystems, aiming to definitively bridge the gap between rapidly advancing technological capabilities and the authentic, complex needs of older adults.

Supplementary information. Not required.

Acknowledgements. The authors would like to thank the residents, relatives and professionals at Vitalia Teatinos, for their kind help during all these experiments.

Declarations

- *Funding:* This work has been funded by the FEDER/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación TED2021-131739B-C21, PDC2022-133597-C42, PID2022-137344OB-C32 projects. It has been also partially funded by by the regional project PHS-2024/PH-HUM-313 (CLINFO-CM) of Madrid (Spain) and the regional grant for research contracts PEJ-2021-AI/SAL-21563.

- *Conflict of interest:* People affiliated to the same university might know about the work (Universidad de Málaga and Universidad Carlos III de Madrid). People working at MetraLabs GmbH might also know about the work. People working at Vitalia Home and, specifically, at Vitalia Teatinos retirement home, might also know about the work. The authors declare no further conflicts of interest.
- *Ethics approval and consent to participate:* The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethical Committee of Vitalia Home. Informed consent was obtained from all subjects involved in the study.
- *Consent for publication:* Not applicable.
- *Data availability:* The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.
- *Materials availability:* Not applicable.
- *Code availability:* Not applicable.
- *Author contribution:* A.J.C., A.I. and J.P.B.R. designed the framework, A.I. designed robot interfaces and HRI procedures. A.T.R. implemented the use cases. A.T.R. monitored and updated the functionality of the robot. A.J.C. conducted the evaluation procedures. A.J.C., A.I. and J.P.B.R. analyzed the collected data. J.P.B.R. coordinated and supervised experiments and technical procedures. J.P.B.R. was in charge of funding acquisition. All authors contributed to writing this manuscript, and all authors have read and approved the final version.

References

[1] WHO, W.H.O.: Ageing and health (2025). https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[2] Feil-Seifer, D., Mataric, M.J.: Defining socially assistive robotics. In: 2005 IEEE C9th Int. Conf. on Rehabilitation Robotics, pp. 465–468 (2005)

[3] Seibt, J., Damholdt, M.F., Vestergaard, C.: Integrative social robotics, value-driven design, and transdisciplinarity. Interaction Studies **21**(1), 111–144 (2020)

[4] Vandemeulebroucke, T., Casterle, B.D., Gastmans, C.: Ethics of socially assistive robots in aged-care settings: a socio-historical contextualisation. J. Med. Ethics **46**, 128–136 (2020)

[5] Norman, D.A., Draper, S.W.: User Centered System Design; New Perspectives on Human-computer Interaction. L. Erlbaum Associates Inc., ??? (1986)

[6] Norman, D.A.: The Psychology of Everyday Things. Basic books, ??? (1988)

[7] Ashley, F.: Accounting for research fatigue in research ethics. Bioethics **35**(3), 270–276 (2021)

[8] Fraune, M.R., Komatsu, T., Preusse, H.R., Langlois, D.K., Au, R.H., Ling, K., Suda, S., Nakamura, K., Tsui, K.M.: Socially facilitative robots for older adults

to alleviate social isolation: A participatory design workshop approach in the us and japan. Frontiers in psychology **13**, 904019 (2022)

[9] Sienkiewicz, B., Radosz-Knawa, Z., Indurkhya, B.: A participatory design approach to using social robots for elderly care. arXiv preprint arXiv:2404.14134 (2024)

[10] Iglesias, A., José, R.V.-A., Perez-Lorenzo, M., Ting, K.L.H., Tudela, A., Marfil, R., Duenas, A., Bandera, J.P.: Towards long term acceptance of socially assistive robots in retirement houses: Use case definition. In: 2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), pp. 134–139 (2020). IEEE

[11] Jerez, A., Iglesias, A., Pérez-Lorenzo, J.M., Tudela, A., Cruces, A., Bandera, J.P.: An user-centered evaluation of two socially assistive robots integrated in a retirement home. International Journal of Social Robotics **16**(9), 2043–2063 (2024)

[12] Marion, O., Stéphanie, R., Dimitri, V., Jean-Gabriel, G.: Contributions of user tests in a living lab in the co-design process of human robot interaction. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 864–869 (2022). IEEE

[13] Fiorini, L., Tabeau, K., D'Onofrio, G., Coviello, L., De Mul, M., Sancarlo, D., Fabbricotti, I., Cavallo, F.: Co-creation of an assistive robot for independent living: lessons learned on robot design. International Journal on Interactive Design and Manufacturing (IJIDeM) **14**(2), 491–502 (2020)

[14] Nertinger, S., Naceri, A., Haddadin, S.: Identifying requirements for the implementation of robot-assisted physical therapy in humanoids: A user-centered design approach. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 2316–2321 (2023). IEEE

[15] Umbrello, S., Capasso, M., Balistreri, M., Pirni, A., Merenda, F.: Value sensitive design to achieve the un sdgs with ai: A case of elderly care robots. Minds and Machines **31**(3), 395–419 (2021)

[16] Gao, Y., Kerse, N., MacDonald, B., Broadbent, E.: Using social robots to enhance cognitive health in older adults with mild cognitive impairment. In: 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 1841–1843 (2025). IEEE

[17] Ghosh, R., Khan, N., Migovich, M., Tate, J.A., Maxwell, C.A., Newhouse, P.A., Scharre, D.W., Tan, A., Mion, L.C., Sarkar, N.: Engaging older adults and staff in the co-design and evaluation of socially assistive robot and virtual reality activities for long-term care: user-centered study. JMIR aging **8**(1), 75288 (2025)

[18] Sanders, L.: An evolving map of design practice and design research. interactions **15**(6), 13–17 (2008)

[19] Veling, L., Villing, R.: Assistive robotics needs for older care: Using authentic citations to bridge the gap between understanding older persons' needs and defining solutions. International Journal of Social Robotics **16**(4), 775–790 (2024)

[20] Abras, C., Maloney-Krichmar, D., Preece, J., *et al.*: User-centered design. Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications **37**(4), 445–456 (2004)

[21] Wilson, R., Keane, I., Jones, R.: Affective responses of older adults to the anthropomorphic genieconnect companion robot during lockdown of the covid19 pandemic. In: 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 1095–1099 (2022). IEEE

[22] Wu, C.-H., Ahtinen, A., Väänänen, K.: Walking outdoor with a zoomorphic mobile robot: Exploration of robot-assisted physical activities for older adults. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 269–276 (2023). IEEE

[23] Olatunji, S.A., Nguyen, V., Cakmak, M., Edsinger, A., Kemp, C.C., Rogers, W.A., Mahajan, H.P.: Immersive participatory design of assistive robots to support older adults. Ergonomics **67**(6), 717–731 (2024)

[24] Ostrowski, A.K., Breazeal, C., Park, H.W.: Long-term co-design guidelines: empowering older adults as co-designers of social robots. In: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), pp. 1165–1172 (2021). IEEE

[25] Juel, W.K., Haarslev, F., Ramirez, E.R., Marchetti, E., Fischer, K., Shaikh, D., Manoonpong, P., Hauch, C., Bodenhagen, L., Krüger, N.: Smooth robot: Design for a novel modular welfare robot. Journal of Intelligent & Robotic Systems **98**(1), 19–37 (2020)

[26] Wengefeld, T., Steffen, Y., Schütz, B., Seichter, D., Scheidig, A., Müller, M.M., Kranz, F., Striebinger, J., Klein, B., Krabbe, M., *et al.*: A mobile robotic assistant to improve social care networking for older and frail people at home: Main results of the morphia-project. International Journal of Social Robotics **17**(9), 1647–1670 (2025)

[27] Krendl, K.A., Broihier, M.: Student responses to computers: A longitudinal study. Journal of Educational Computing Research **8**(2), 215–227 (1992)

[28] Koch, M., Luck, K., Schwarzer, J., Draheim, S.: The novelty effect in large display deployments–experiences and lessons-learned for evaluating prototypes.

In: Proceedings of 16th European Conference on Computer-supported Cooperative Work-exploratory Papers (2018). European Society for Socially Embedded Technologies (EUSSET)

[29] Smallman, M., O'Donovan, C.: Co creation in social innovation. Edward Elgar Publishing (2023)

[30] Branch, R.M., Varank, İ.: Instructional Design: The ADDIE Approach vol. 722. Springer, ??? (2009)

[31] Foster, E., Towle Jr, B.: Software Engineering: a Methodical Approach. Auerbach Publications, ??? (2021)

[32] Bouchetara, M., Amrani, A.F.Z., Bedaida, I.E.: The implementation of a quality management system in accordance with iso 9001: 2015 standard: A case study. International Journal of Economics & Business Administration (IJEBA) **10**(1), 261–286 (2022)

[33] Meinel, C., Leifer, L., Plattner, H.: Design Thinking: Understand-improve-apply. Springer, ??? (2011)

[34] Lobo, A., Ezquerra, J., Gómez Burgada, F., Sala, J., Seva Díaz, A.: Cognocitive mini-test (a simple practical test to detect intellectual changes in medical patients). Actas luso-espanolas de neurologia, psiquiatria y ciencias afines **7**(3), 189–202 (1979)

[35] Lindsay, S., Jackson, D., Schofield, G., Olivier, P.: Engaging older people using participatory design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1199–1208 (2012)

[36] Barrett, J., Herriotts, P.: Running focus groups with older participants. Focus Groups: Supporting Effective Product Development. Taylor and Francis, London, 115–144 (2003)

[37] Iglesias, A., García, J., Garcia-Olaya, A., Fuentetaja, R., Fernández, F., Romero-Garcés, A., Marfil, R., Bandera, A., Ting, K.L.H., Voilmy, D., *et al.*: Extending the evaluation of social assistive robots with accessibility indicators: the ausus evaluation framework. IEEE Transactions on Human-Machine Systems **51**(6), 601–612 (2021)

[38] Iglesias, A., Viciana, R., Pérez-Lorenzo, J.M., Ting, K.L.H., Tudela, A., Marfil, R., Qbilat, M., Hurtado, A., Jerez, A., Bandera, J.P.: The town crier: a use-case design and implementation for a socially assistive robot in retirement homes. Robotics **13**(4), 61 (2024)

[39] Weiss, A., Bernhaupt, R., Lankes, M., Tscheligi, M.: The usus evaluation framework for human-robot interaction. In: AISB2009: Proceedings of the Symposium

on New Frontiers in Human-robot Interaction, vol. 4, pp. 11–26 (2009)

[40] Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view1. MIS quarterly **27**(3), 425–478 (2003)