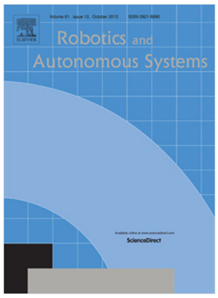
Journal Pre-proof



User feedback and remote supervision for assisted living with mobile robots: A field study in long-term autonomy

Matteo Luperto, Marta Romeo, Javier Monroy, Jennifer Renoux, Alessandro Vuono, Francisco-Angel Moreno, Javier Gonzalez-Jimenez, Nicola Basilico, N. Alberto Borghese

PII: S0921-8890(22)00096-3
DOI: https://doi.org/10.1016/j.robot.2022.104170
Reference: ROBOT 104170

To appear in: *Robotics and Autonomous Systems*

Highlights

- Socially assistive robots are promising tools for older adults independent living
- Private apartments are challenging environments, impacting the robustness of robots
- Results from a 130 weeks study in 10 older adults apartments using autonomous robots
- Remote supervision increased the acceptance of socially assistive robots
- Lessons learned from managing a long-term deployment, towards long-term autonomy

Noname manuscript No.
(will be inserted by the editor)

User Feedback and Remote Supervision for Assisted Living with Mobile Robots: a Field Study in Long-Term Autonomy

Matteo Luperto*[1] · Marta Romeo[2] · Javier Monroy[3] · Jennifer Renoux[4] · Alessandro Vuono[1] · Francisco-Angel Moreno[3] · Javier Gonzalez-Jimenez[3] · Nicola Basilico[1] · N. Alberto Borghese[1]

Received: date / Accepted: date

Abstract In an ageing society, the at-home use of Socially Assistive Robots (SARs) could provide remote monitoring of their users' well-being, together with physical and psychological support. However, private home environments are particularly challenging for SARs, due to their unstructured and dynamic nature which often contributes to robots' failures. For this reason, even though several prototypes of SARs for elderly care have been developed, their commercialization and widespread at-home use are yet to be effective. In this paper, we analyze how including the end users' feedback impacts the SARs reliability and acceptance. To do so, we introduce a Monitoring and Logging System (MLS) for remote supervision, which increases the explainability of SAR-based systems deployed in older adults' apartments, while also allowing the exchange of feedback between caregivers, technicians, and older adults.We then present an extensive field study showing how long-term deployment of autonomous SARs can be accomplished by relying on such a feedback loop to address any potential issue. To this end, we provide the results obtained in a 130-week long study where autonomous SARs were deployed in the apartments of 10 older adults, with the aim of possibly serving and assisting future practitioners, with the knowledge collected from this extensive experimental campaign, to fill the gap that currently exists for the widespread adoption of SARs.

Keywords Socially Assistive Robots · Long-Term Autonomy · Field Study

We gratefully acknowledge the European Commission H2020 project MoveCare grant ICT-26-2016b – GA 732158 and Essence SC1-PHE-CORONAVIRUS-2020-2B - GA 101016112 for financial partial support to the authors of this work.

*Corresponding Author: M.Luperto
E-mail: matteo.luperto@unimi.it
[1]Applied Intelligent System Lab, Department of Computer Science, University of Milan, Milano, Italy
[2]University of Manchester, Manchester, UK
[3]Machine Perception and Intelligent Robotics Group (MAPIR), Department of System Engineering and Automation, University of Malaga, and Instituto de Investigación Biomédica de Málaga–IBIMA, Spain
[4]Machine Perception and Interaction Lab, Örebro University, Örebro, Sweden

1 Introduction

An ageing population presents many challenges to the healthcare and social systems of societies. Assistive technologies are often investigated as a way to tackle those challenges due to their high potential in providing innovative and cost-effective improvements to the quality of elderly care [5]. At-home assistive solutions, in particular, aim at sustaining the independent living of older adults by remotely monitoring their health and well-being, pursuing to postpone as much as possible their admittance into a more controlled (but often more expensive and overpopulated) care facility.

An emerging instance of this approach is represented by Socially Assistive Robots (SARs), a technology often integrated with Ambient Assisted Living (AAL) settings. SARs can perform remote monitoring, stimulation, and assistance while facilitating communications between the users and the caregivers [9,10]. However, although considerable research effort has been devoted to the engineering

Fig. 1: A Giraff-X robot inside the house of an older adult.

and testing of SAR-based solutions [18, 13], widely deployable systems are not sufficiently mature.

This is mainly caused by flaws in terms of *reliability* and *robustness*, especially in home environments, which are particularly challenging for robots due to their unstructured and dynamic nature [21], as can be appreciated in Fig. 1. Difficulties, such as intricate corridors, narrow doorways, or movable objects, degrade the robot's *performance* or even cause *failures* that typically require costly recovery procedures, often involving external interventions from trained personnel. As a consequence, the safety concerns that arise in their potential users and the lack of trust towards the robot have a negative impact on SARs acceptance [42]. In this sense, one critical issue to be avoided is the users' feeling of not being in control, experienced when they do not understand the robot's behaviour.

In projects involving SAR deployments over a long-term time interval, great effort is spent to ensure the robustness of the system and, particularly, that of the SAR. However, the list of potentially troublesome situations is so broad that forecasting and eliciting them with laboratory tests becomes unfeasible. Consequently, *control* and *remote supervision* become key functionalities to integrate into the system [11]. From one side, this increases the costs of system management, but from the other it offers the opportunity to mitigate the effects of unexpected behaviours or failures on user's acceptability: if the robot showed an anomalous behaviour, end-users feel safer knowing that there is someone monitoring the situation. Furthermore, they would be able to receive *explanations* about the reasons for the perceived anomaly. It must be stressed that, sometimes, even the intended robot behaviour can be perceived as anomalous by users if robot actions are not understood.

In this paper, we present an extensive field study showing how long-term autonomous SAR deployment can be accomplished by addressing the aforementioned issues. To this end, we provide the results obtained in a data collection of more than 130 weeks in apartments of 10 older adults during a pilot study for the H2020 MoveCare project [26], with the aim of enlightening the technological requirements needed for the long-term autonomy of SARs.

At first, we show how the engineering, testing, and refinement of the robotic platform resulted into achieving robust core functionalities with an adequate level of performance. Then, we describe a web-based cloud-robotics Monitoring and Logging System (MLS) designed to provide remote supervision by enabling supervisors and caregivers to: monitor the correct functioning of each deployed SAR-based system, detect anomalous robot behaviours, provide an understanding of the chain of events that lead to the anomaly, and fix the problem. The MLS, combined with the users' capability of signalling system strange behaviours and asking for an explanation from the technicians, was used to create a feedback loop between caregivers, technicians, and older adults. The latter showed that the system *explainability* increased from the user's perspective, while providing technicians/caregivers with a way to both reassure the users and improve the system by adapting it to their needs. As a result, the use of the MLS during the pilot study increased the acceptance of the SAR-based system in the event of failures and anomalies.

Finally, we provide detailed results of the insights and findings obtained in the MoveCare project by analysing and discussing the critical issues that arose during the long-term deployment of the SARs, the countermeasures that were taken, and the robot evaluation provided by users. The final result shows a system with adequate performance and whose confidence and acceptance increase by addressing failures through explainability. We hope that the knowledge collected from this extensive experimental campaign could serve to fill the gap that currently exists for the widespread adoption of SARs and to assist future practitioners in the field in the development of new and improved SARs.

A preliminary version of this article has been presented in [32], where we describe the MLS. In this paper, we provide a substantially improved evaluation, discussion, and report of the lessons learned from the long-term experimental campaign. A detailed description of the functionalities provided to users by our platform (e.g., monitoring and support) is beyond the scope of this paper and is described in [29, 26].

2 Related Work

The term SAR refers to a class of robots defined as being at the intersection of socially interactive robots, focusing on engaging and *stimulating* the users through social and non-physical interactions, and assistive robots, whose aim is to overcome the physical limitations of the patients by helping them in their daily activities (such as getting out of bed, brushing teeth, walking, etc.)[10].

SAR-based systems represent a promising technology aiming to provide assistance to older adults through social interactions, and their functionalities are often investigated with structured interviews and controlled pilot studies, as in [15, 9, 37]. The work of [7] presents an overview of several SAR platforms developed in the last twenty years, showing their evolution from early prototypes, tested in laboratories and controlled environments, to technologically ready systems to be deployed in heterogeneous contexts with their end-users.

In this context, SARs are often based on autonomous mobile platforms that, thanks to a high level of integration into AAL environments and cloud-based frameworks [11], are able to provide services such as the delivery of messages and reminders, teleconference with family members or caregivers, and guidance through physical or cognitive exercises [13]. However, in the aforementioned works, the cloud-based infrastructure was not acting as a monitoring system, and it was not used as a way to provide explainability to the end-users. It was more an instrument for the researchers to be present during the at-home deployment of the SAR-based system.

In recent years, several projects have investigated the deployment and use of SARs in home environments [18, 17, 13]. While these works have successfully assessed the use of SARs for a short period of time, there is still a significant gap between the current abilities of SARs and those required for a robot to be able to reach long-term autonomy in uncontrolled environments [21]. As a result, experiments involving SARs for fostering independent living of older adults often offer in-presence supervision of the robot or investigate the possibility of creating a mutual care network where the robot and the user support each other [22].

Other works investigated several aspects related to the long-term deployment of SARs. As an example, the CARESSES project studied how the user's social background influences the interactions with a robot [36]. The followed approach was built on the rationale that robot's skills integrating awareness of user's culture while being capable of interpreting and conveying social signals would boost acceptability. Feasibility and acceptability of the long-term use of a telepresence robot within the house of elders was also studied in [12, 6]. In such settings, however, the robot was remotely controlled by a caregiver and not able to navigate autonomously.

Recent works like [20] or [39, 8] have done a remarkable effort towards the long-term and widespread adoption of SARs by deploying them for several weeks in settings such as assisted living facilities. Other relevant long-term applications of autonomous social robots are those of [33] and [4]. Unlike our setting, in both studies, the robot was deployed in less challenging and more controlled environments. In our context, our aim is to provide an assistance robot to serve older adults at their own home.

A similar system to the one we are proposing can be found in the work done for the CompanionAble, SERROGA, and SYMPARTNER projects [19, 41, 17, 18]. In their works, similarly to what we discuss in Section 6, authors presented results on the performance of a SAR-based system deployed in private apartments. Another recent research effort presented Hobbit, a service robot focused on fall detection, while also offering additional services such as reminders and entertainment suggestions [13]. In [3], this robot was used for three weeks in the households of 18 users.

The EnrichMe project [39, 8] assessed the feasibility of long-term deployments within the house of 10 elders for 10 weeks. The main objective of this project was to provide tools and applications that are of everyday use to assist users at home. These tools focused on health monitoring, complementary care, and daily support.

With respect to the above studies, our work distinguishes itself along different features. First, we provide a significantly longer experimental campaign, where multiple robots were used at the same time for several months at the end-users' own houses. Moreover, the behaviour of our robot is mostly proactive, as it is the robot that initiates scenarios by moving and interacting with the users. This is a far more challenging scenario, especially in a long-term deployment, as the robot needs to search for the users (who are performing their regular

Project	Pilot		Scenarios		Search for the User	IoT Integration	Remote Supervision	Autonomous Navigation
	Users	Duration	Init. Robot	Init. User				
Hobbit [13,3]	18	3 w	✓	✓	✓	-	-	✓
SYMPARTNER [18]	20	1 w	✓	-	✓	-	✓	✓
SERROGA [17]	9	1 w	✓	-	✓	-	✓	✓
CompanionAble [41,19]	15	2 d	✓	✓	✓	-	✓	✓
EnrichMe [8,39]	11	10 w	-	✓	-	✓	-	✓
Fiorini et al. [12]	20	2 w	✓	-	-	-	✓	-
Giraffplus [6]	15	up to 1 y	✓	-	-	✓	✓	-
Our work [1,29]	13	10-16 w	✓	✓	✓	✓	✓	✓

Table 1: Overview of different works involving a long-term use of a SAR in private apartments of older adults, compared to our work for the MoveCare project. The table distinguishes robots that were proactive and started the interaction with the user inside the apartment (Init. Robot) and those where it is the user who starts the interaction with the robot (Init. User). While some robots assume that the location of the user is known when the scenario starts, some robots are required to search for the user in the entire apartment (Search for the User).

daily activities) by navigating autonomously in the apartment, as it does not know their exact location in advance. As an example, the Hobbit robot [13, 3] is called exclusively upon the user's request by using a button associated with a target location in the environment, while in [8], similarly, is the user who starts to interact with the robot by pressing a button on the robot's GUI. Finally, all the deployments described in the aforementioned works do not explicitly discuss how to remotely control the SAR-based systems, nor give the opportunity to caregivers and technicians to easily engage with them. A general overview of these differences is provided in Table 1 where we list the main projects that involved long-term deployments of SARs in private apartments, highlighting the differences between settings along key dimensions related to the role of the robot.

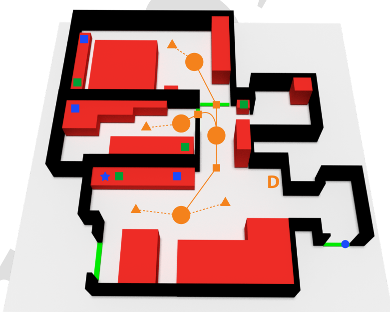


Fig. 2: An example of the setup in one of the pilot houses. The robot uses a *topological map* (in orange), represented as a graph where nodes represent rooms (circles), and edges represent a connection in the map. Each room could be associated with a set of *poses* (triangles), location of interests for the robot, where it could identify the user. Poses are the targets of robot navigation. When idle, the robot is at the docking station (D). The IoT network (blue) monitors the user with PIRS (squares) and a door sensor (circle). A concentrator (star) provides connectivity to the cloud. Smart microphones (green) listen to voice commands from the user.

3 System Architecture

The MoveCare system is an AAL framework developed around Giraff-X, an autonomous mobile SAR, and is composed of several components installed in the apartments of older adults living independently on their own [26]. Besides the robot, the other components are an Internet of Things (IoT) network, which is used to monitor the user and to provide information to the other components, a Community Based Activity Center (CBAC) [28], and a Virtual Caregiver (VC), a software component acting as platform orchestrator [38].

All components interact to carry out a set of *scenarios* designed to provide assistance, support, and both physical and cognitive stimulation to older adults. Examples of such scenarios are described in [31,43] and shown in demonstrative videos[1].

More precisely, the IoT network collects data on the users' behaviours for monitoring purposes. It is composed of both passive sensors (e.g., PIRs to identify the user activity in a given room) that

[1] http://www.movecare-project.eu/index.php/results/videos/

do not require user interaction, and active sensors (e.g., a weight scale and a sensorized ink pen) that collect data upon direct usage. A set of microphones is deployed in each room to give commands to the system, and to detect requests for help from the users in case of emergency. The CBAC provides a set of digital activities for social, cognitive, and physical stimulation while also supporting monitoring of the users and the possibility to provide notifications. It could be accessed through a tablet (provided by the project) or through a TV set-top box connected to the main TV of the users. The digital activities of the CBAC are also integrated with active sensors that can be used to play.

An example of how these components are installed in a target environment is shown in Fig. 2. Full details of the components, the functionalities, and the architecture of the system are beyond the scope of this work and can be found in [29].

3.1 Giraff-X Socially Assistive Robot

Giraff-X, an enhanced and autonomous version of the Giraff teleoperated robot explicitly developed for AAL[35,6], is the main actor of the system and is responsible for the execution of the scenarios.

The robot is designed to operate and move unsupervised for a continuous period of time in older adults' private apartments. See Fig. 3 for a detailed view and [30] for more details about the robot setup. This robot conveys a set of functionalities in the form of *interventions* embodying the caregiver in the users' house and following these steps:

1. receives the intervention to be executed via MQTT messages;
2. undocks (if necessary);
3. safely navigates to the expected user location (updated by the system in real-time thanks to the IoT network);
4. locates the user or, if not found, performs a search in the whole house (Section 3.1.3);
5. approaches the user, taking into consideration the proxemic distances [14];
6. interacts with the user to carry out the specified scenario (Section 3.1.4);
7. provides feedback to the VC;
8. returns to the docking station (if there is not any other intervention planned in a short time interval [16]).

During the day, the robot is waiting at its docking station for an intervention to be requested. Interventions can be triggered by the system (when

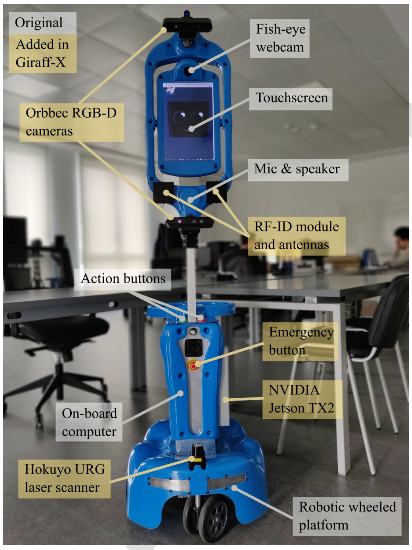


Fig. 3: The Giraff-X mobile robot.

there is a need to interact with the user) or can be directly requested by the users, in the form of assistance services (e.g., in case they need help). In addition, some interventions are directly triggered by the robot itself, as part of its self-management, aiming to maintain a proper autonomy level (e.g., triggering auto-docking if the battery level is critical). A full list of the possible robot's interventions is shown in Table 2. On average, the robot performs 2-5 interventions per day, the main purpose of the robot being to perform these interventions autonomously, without any external help. In this context, robustness is crucial, as a failure of the robot may undermine the perceived utility in the eyes of the users.

To carry out an intervention, the robot is provided with a set of core functionalities or behaviours, namely autonomous navigation and mobility, self-management, and Human-Robot Interaction (HRI).

3.1.1 Autonomous Navigation and Mobility

Having a safe and robust navigation within the users' home is of paramount importance, as the robot is required to efficiently search, approach, and interact with the user while avoiding collisions and/or navigation failures.

Navigation is performed on a 2D static map of the apartment, built at installation time [30],

Intervention	Label	Description	Ref.
Reminder	REM	The robot reminds the user to perform one of the activities scheduled by the framework for monitoring or stimulation purposes.	[24, 23, 43, 28]
Spot Question	SQ	The user is asked to answer questions to monitor their cognitive state.	[38]
Weight Measure	WM	The robot asks the user to measure the weight on a connected weight scale. After the user complies, the robot gives an acknowledgment to the user that the data is collected.	-
Neuropsy. Tests	NPS	The user is asked to perform a cognitive assessment using digitalised tests on the tablet under the guidance and supervision of the robot.	[25, 31, 27]
Call for Help	CH	The user requests assistance by voice, which is detected by an environmental microphone; the robot searches for the user and confirms the request; a phone call alerts a caregiver; the caregiver remotely teleoperates the robot.	[29]
Find Lost Object	OB	The user requests the robot to help find an object to which an RFID tag has been previously attached, searching around the house.	[29]
Teleoperation	TO	A caregiver request permission to teleoperate the robot, which must be granted by the user. A web-based interface is provided to the teleoperator to easily and safely control the robot.	[29]
Go Dock	DS	Either the user (via voice command), the system, or the self-management module within the SAR, requests the robot to go to its docking station and stay there.	-
Talk	SAY	The robot reproduces a voice and text message to inform the user.	-

Table 2: List of interventions performed by the robot.

and enriched with the locations of key points (e.g., the position of the docking station and of the environmental sensors; see Fig. 2 and [29] for more details). Among these key points, a set of *poses* is given for each room by manually inserting them in the map at the set-up time. They constitute targets in robot navigation, also representing locations of interest where the robot can start performing an action (e.g., searching the user). As an example, poses can be placed at locations where the user is likely to be found (close to the sofa, a table, or next to the bed; see Fig. 2). The robot moves from one key point to another using the navigation functionalities offered by the ROS navigation stack[2]. To enhance robustness, the 3D point-clouds from two RGB-D cameras are used for obstacle avoidance, as well as to detect the user in a wide variety of situations [14].

Despite all the sensors and algorithms employed to ensure a safe and robust navigation, the robot must also be able to handle failures and try to avoid them in future occasions. To cope with challenges of older adults apartments, such as narrow doorways or corridors, we introduced a navigation assistant (described in [34]) that detects and autonomously inserts waypoints within the problematic areas.

[2] http://wiki.ros.org/navigation

3.1.2 Self Management

During a long-term deployment, a robot should be able to maintain itself operative and self-evaluate its status, avoiding that errors and failures accumulate over time until the system stops working.

For that, battery management is crucial. Thus, we have provided the robot with the ability to automatically perform a docking manoeuvre, based on observing a visual landmark placed on top of the docking station (unlike navigation, which is based on data from a 2D lidar). Besides, the robot continuously updates its status to the VC, so that the system is always aware of the robot situation and can decide when to send the robot to recharge. In situations where the robot cannot communicate with other components (e.g., lack of connectivity) the robot implements its own offline mechanisms to ensure that the battery level remains above a healthy threshold.

Another critical aspect for an autonomous mobile robot is its localisation within the environment. Despite the robustness and recovery mechanisms in place for state-of-the-art localisation algorithms, over a long-term operation localisation errors tend to accumulate and performance to degrade, eventually jeopardising the navigation. To reduce this risk, we reset the robot internal localisation estimation every time the robot reaches the known location of the docking station.

When the robot is unable to complete its task autonomously (e.g., a door is not fully opened or a chair is excessively reducing the free space for navigation), Giraff-X informs the VC about the possible cause of the problem and automatically returns to its docking station. However, should the problem arise when returning to the docking station, this issue becomes critical. In these circumstances, and after three failed docking attempts, the robot asks for help from the user (via voice). Giraff-X then instructs the user to assist it with the problem (e.g., by removing the object that interferes with navigation), manually move the robot to its docking station, or contact a technician. If the user does not respond, the robot executes its recovery behaviours until a critical battery level is reached (< 10%). Then, to preserve the battery, it automatically shuts down.

3.1.3 Search for the User

Searching for a user within an apartment is a challenging task because the user location is not known in advance and the user could move across different rooms while the robot is performing the search. When an intervention request is received, the robot starts moving from its position (typically its docking station) and performs a full search of the house until either the user is found, the intervention is cancelled (by the system or by the user), or a timeout is reached. The search is performed room after room, with the robot reaching all poses within a target room and doing a 360° turn in each pose trying to locate the user. The robot can detect the user also while navigating from pose to pose.

The robot is constantly informed by the VC of the estimated room where the user could be located (see Section 3.2 and [38]) using data coming from PIR sensors. However useful, this information could be inaccurate, as the user may be moving back and forth between different rooms (thus triggering different PIRs) or may stay still for a long time, and hence not being detected by the motion sensors (e.g., sitting when working at a table).

When the search is started, the robot goes to the last estimated location. If a new estimated location is sent by the VC in the meantime, the robot interrupts its search and starts moving towards the new room. If the user is not found in the expected room, the robot performs a full search of all rooms in the house (following a ordered list of rooms sent by the VC). As a result of this behaviour, the robot can search for the user in the same room multiple times (if it is signaled that the user is in that room by the VC while the robot is searching in or navigating towards another room). If the user is already in front of the robot when the search is started (e.g., from a previous intervention) no search is performed. During the user search procedure, the user is not instructed to wait or help the robot (e.g., by going towards the robot), but they are told to continue with their activity.

After a successful search, i.e. when the user is detected by the robot, the search is ended and the robot starts the next behaviour (i.e. the user approach [14]).

3.1.4 Human Robot Interaction

The HRI functionality of the robot is in charge of managing and facilitating the interaction between the robot and the older adults. It is composed of two functional modules: a Dialog Manager (DM) and a Speech Module based on state-of-the-art cloud services. (The technical details can be found in [31,29]).

The main communication mechanism of the Giraff-X is voice. Giraff-X's portrait screen was used to display subtitles for what the robot said and to inform users when the robot was listening to their answers. Giraff-X's screen was also used to display a friendly robotic face at all times, to make the platform more appealing to the users. We decided to base our Speech Module on two state-of-the-art speech services because of their robustness and continuous update. However, to mitigate problems due to internet connection latency or environmental noise, we decided to allow users to have the possibility of answering the robot using the red and green buttons that were already part of the hardware of the platform. The green button can be used for positive answers and the red button for negative answers.

Whenever the robot approached the user for an intervention, the HRI module first told the user what they were supposed to do and asked if they wanted to complete the intervention at that moment. If the user answered positively, they would go on with the intervention. If the user answered negatively, the robot labelled their answer as "Later" and informed the VC to reschedule the intervention. If, at any point during an interaction, the module did not receive or was unable to process an answer from the user for 30 seconds, it would ask the same question again; if still no answer was received, the interaction would stop and be considered unsuccessful.

3.1.5 Robot-System Integration

Giraff-X interacts with the other system components through JSON messages, sent through an MQTT channel to a shared cloud platform. This way, the robot receives the intervention to execute for a specific scenario and, upon completion, provides a report including information regarding the steps carried out for its completion and the outcome of the execution (success/failure). Besides this, the robot constantly updates the system with a message reporting its status: its pose in the map, its topological location, its battery status, its current and pending tasks, and a list of objects/people detected (using vision).

3.2 IoT and Orchestrator

The IoT network provides data to both the robot and the VC [26]. It is made up of a *concentrator*, providing connectivity and internet access to all components, an IoT *sensors network* of PIRs (to detect and track the location of the user), a door contact sensor that detects the user entering/exiting the house, and a set of *smart microphones*, whose range covers the entire apartment and which are used to detect pre-defined commands for the system (i.e., a request for help). IoT data is analysed by the VC, a cloud software component, which is the system orchestrator that controls, through MQTT, the execution of the scenarios [38].

Within the VC, the user-location module gathers data from the IoT sensor network to detect if the user is available (i.e., in the home) as well as their room location. The robot then uses this information to reach the user as quickly as possible. In addition, the VC decides if and when the robot should perform an intervention based on a set of constraints described in [38, 29]. The two most important constraints are that no intervention should be issued if the user is not at home (user-at-home constraint) or during the night (night-time constraint). Indeed, allowing the robot to move in these contexts is not only undesirable from the user's perspective, but also presents a higher risk of robot failure. The first constraint (user-at-home) is validated using the user-location module; users are considered as "Outdoor" if they are not detected in any of the rooms through the IoT sensor network. The second constraint (night-time) is validated through ad-hoc time schedules, decided with the user feedback and reflecting when they

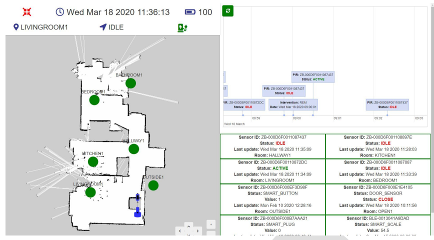


Fig. 4: The MLS interface shows all events in real-time and provides the position and status of both the robot and the rest of the components. On the left: on top of the robot map (showing the robot's current location), there is the robot status (in this example: idle at full charge at the docking station). On the right: a list of all events in a timeline (top) and the status of all IoT sensors (bottom).

considered that it was acceptable for the robot to act. Following this constraint, the VC schedules interventions only between 9:00 AM and 19:00 PM.

For robustness reasons and in order to prevent corner cases, we implemented redundancy between the VC and the robot. Before issuing an intervention, the VC ensures that all the constraints are respected, including, but not limited to, the user-away and night-time constraints. Upon receiving an intervention and before starting to act, the robot validates the user-away and night-time constraints. This robot constraint prevents the robot from performing an intervention between 22:00 AM and 8:00 AM. The only exception to these constraints is the call for help scenario (CH - see Table 2), which must be triggered regardless of the user's detection status or time of the day for safety reasons. Note how, while necessary, the constraints on the intervention execution reduce the time windows available for the robot to carry out the interventions: the robot is allowed to move only during the daytime and while the user is at home.

4 Monitoring and Logging System

In order to provide efficient remote supervision of SARs, a remote monitoring system should be able to:

- provide a real-time remote overview of the robot status and other components of interest;
- replay and inspect past robot (and system) action flows;
- directly access to the robot platform;

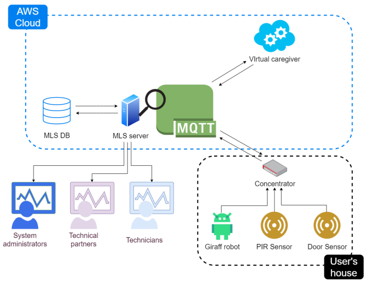


Fig. 5: The MLS architecture is developed as a cloud-based platform which receives data from the robot and the sensors via MQTT messages.

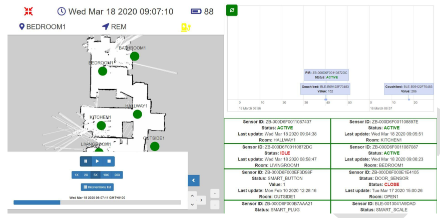


Fig. 6: The replay functionality is used to better understand a particular behaviour by tracing back all events that lead to it.

– support multi-robot management.

To meet these needs and to provide supervision and control for the long-term deployment of the MoveCare system, a cloud-based MLS was developed. This MLS could be accessed by all authorised users through a web interface designed to be easily used by caregivers or technicians without experience in robotics, as well as experienced robotics researchers. It allowed technicians and caregivers to inspect, in real-time, past and current events happening at users' homes, to provide *remote support* to older adults. Besides, it could be used to transparently assess the proper functioning of a single installed system, or even multiple systems at-a-glance, including monitoring anomalies and failures (e.g., a faulty IoT sensor). Not only that, thanks to the MLS, users receive feedback after signalling an anomalous behaviour or asking for an explanation on some unexpected events (e.g., "the robot approached me and talked to me this morning but, after that, it remained static in its position"). In this scenario, the system could be used to monitor the reported event, to understand its causes, and to ultimately provide assistance and explanations to the users. Due to all of this, the MLS allowed us to remotely handle most of the issues that arose during the pilot experimental campaign, effectively reducing the number of maintenance interventions on site.

Fig. 4 shows the main interface of the MLS, which can be accessed from any web browser and provides a rich view of the current state of the system for a selected pilot user. As can be seen, through the MLS, system administrators and technicians were able to monitor:

1. the status of all IoT components;
2. the map of the environment and the real-time position of the robot within it;
3. the status of the robot (its pose, the percentage of charge and voltage supplied to the battery, its current goal location, how navigation progresses, etc.)
4. the status of the VC (current scenario and estimated topological position of the user inferred from the IoT environmental sensors)

To do that, the MLS has been developed as a cloud-based architecture, shown in Fig. 5, that receives MQTT messages from the system components, updates their status (and its associated user), transmits the updated information in real-time to all other components monitoring the same user, and finally stores the events into a database.

In addition, every 30 minutes, the MLS saves a snapshot of the current status of all the installed systems in its database, so that system administrators have the possibility to *replay* the history of events on an interactive timeline (as in Fig. 6). This replay can be configured by specifying the interval to inspect, the desired playback time and speed (from 1x to 20x), and also the resolution of the temporal axis to refine/coarse the level of historical detail.

This feature provides key support for system failures, as it easily allows the retracing of the chain of events that led to the anomaly in chronological order. Finally, the proposed web architecture allows accessibility and flexibility to the SAR-based system, as any other types of sensor/data can be easily incorporated into the system, provided that they can be integrated inside the MQTT communication infrastructure.

5 Prepiloting the System

In order to ensure a proper functioning of the system, extensive testing of all single components has been performed at first in the laboratory. Repeated lab tests allowed us to prepare configurations capable of satisfactory performance: navigation, autonomous docking, and speech recognition trials obtained success rates well above 90% for diverse in-lab conditions. The controlled setting of a laboratory, however, cannot fully capture those real-world complexities that have a remarkable impact on the system, especially within domestic environments. For this reason, in-lab tests have been complemented with two prepilot experimental campaigns. The additional evidence gathered in the field allowed us to solve new issues and improve the system when dealing with more complex contingencies.

5.1 Preliminary Controlled Tests in a Private Apartment

The first prepilot was conducted in Milan (Italy) with a 9-days trial inside an apartment with $\sim$65 m^2 of free area and 3 fully furnished rooms with cluttered spaces. To align with the pilot requirements, a single user lived in the apartment throughout the duration of the test. We decided to stress the system as much as possible in order to gather significant data. For that, the number of interventions generated during the 9 days was comparable to what would have been expected during a month of normal system usage. Concretely, we performed 116 robot interventions in which the robot needed to find, detect and approach the user to provide a voice message. Full details are provided in [30].

In this stage, the robot achieved an acceptable performance in all the above sub-tasks, so that only 3 interventions resulted in failures, requiring manual recovery. However, the experiment revealed problematic situations that were not identified in the in-lab tests. The robot often placed itself in positions very close to the user in order to facilitate the HRI. However, due to the cluttered environment and the typical locations where the user was found (e.g., sitting at a table or on the couch), such positions required difficult manoeuvres for the robot to return to its docking station, resulting in 22 failed docking attempts. To tackle this problem, a more robust behaviour was implemented, in which additional docking attempts were repeated in case of failure.

Fig. 7: The robot dealing with a narrow passage, a door that can be only partially opened due to furniture behind it.

Another critical issue that emerged from this preliminary campaign involved narrow passages characterised by having less than 70cm of free space (see Fig. 7). This reduced margin for manoeuvring significantly slowed navigation in several occasions. Although such a problem had been observed in the lab before, its likelihood in the field was severely underestimated. After this preliminary campaign, the navigation module was improved in order to generate smoother and more cautious trajectories in the proximity of passages [30].

In all the above situations, the use of the MLS proved to be fundamental in reconstructing the chain of events that led to failures, hence providing the key ability of identifying the actual reasons behind the observed issues.

5.2 Preliminary Testing with End-Users in a Care Facility

One additional step needed to ensure the robustness of the system was to test it with potential end-users. To do so, an additional prepilot campaign was conducted in a care facility in Extremadura (Spain) with the aim of assessing if the system was robust and user-friendly enough for our target users. Thus, one of the apartments in the facility was set up with an installation of the entire system, and 7 residents were asked to individually spare some time there to test the different functionalities within manually triggered scenarios that involved the robot (Fig. 8).

Then we collected the users' satisfaction and robot acceptability through the same questionnaires

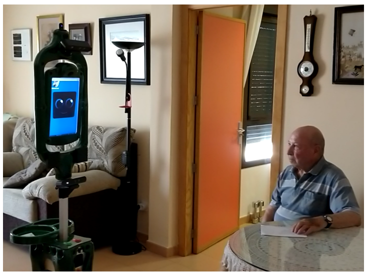


Fig. 8: A resident of the care facility in Spain participating in the prepilot experimentation.

that we later used for the main field experimentation. The 7 participants consistently reported high scores for the robot, both in terms of feeling safe during its navigation/approach and of the ease of interacting with it when needed. Testing the system in this context gave us the possibility to include the feedback of potential users in the final steps of development of the whole platform. Specifically, we improved the navigation module by better calibrating the robot's speed as emerged from suggestions given by the residents. Some scenarios were also modified to make them simpler to execute.

6 Field Experimentation

In this section, we start by describing the setting of our field experimentation. We then present a series of results in terms of the quantitative performance of the robot, the qualitative evaluation of selected case studies, and the system's acceptability assessment.

6.1 Setup of the Pilot Experiment

To be eligible for our study, participants had to be over 65, not suffering from any cognitive impairment, and live independently on their own. To provide well-suited conditions for robot navigation, the apartments had to satisfy some selection criteria too. For example, there could not be any pets, rugs, or stairs. A total of 14 older adults, with an average age of 76.86, were selected from two different regions: 7 were residing in Milan, Italy (we shall refer to this group as `ITA`), while the other 7 were from Badajoz, Spain (group `ESP`). In the `ITA` group, 3 participants lived in apartments hosted by an Assistive Living (AL) facility (where independent living is supported by sporadic assistance). We report data extracted from 10 of the 14 total participants of the field study, with an average age of 75.3, as discussed in Section 6.2.

The experimental campaign was made up of two rounds, the first taking place from September to December 2019, while the second from January to April 2020. Our SAR fleet was active for 132 weeks in total. Some users participated in both rounds (some first-round participants decided to keep using the system for the second round too). This, combined with the users' different availabilities, resulted in each of them participating for different periods of time. However, each participant included in the results presented in this paper accrued at least 10 weeks of system usage.

During the installation of the system, the robot acquired a map of the apartment by exploring only the rooms where it was supposed to operate: as discussed with the Ethics Committee that approved our study, the robots were not allowed to map or enter the bathrooms for privacy and safety reasons. Moreover, at this stage, the users could request not to map some additional rooms (e.g., their bedroom). Once the installation was completed, the users attended a training session where they engaged in 2-4 interventions with the robot, under the supervision of a caregiver. During this session, the users were also briefly introduced to the MLS and got to understand which parts of the system were kept under supervision by it.

For the entire duration of the pilot, regardless of night or day, the robots were "on" (either moving or idle), even if the users were away for a few days. During the pilot, users had the possibility to disable the system in many ways. They could use a wireless button to disable/enable the entire system via software. In addition, they could push the emergency stop button placed on the robot. The emergency button triggered a hardware disconnection of the motors of the robot from the power supply and a software block of the robot capabilities (e.g., to talk). This way, the robot was unable to move autonomously and could be manually moved by the users. To restore it, users had to manually pull the emergency button. Lastly, users could always request remote support from a technician to have the robot remotely turned off (e.g., when they went away for a few days). The technician could also provide assistance in case of (actual or perceived) unexpected events by retracing recent system activity through the MLS.

At the end of the pilot, the system was uninstalled. Approximately a week later (upon availability of the user) a set of questionnaires evaluating the experience were administered by a member of the team who did not interacted with the user nor took part in the robot pilot setup and management. To evaluate the role and the acceptability of the robot within our platform, we developed, following the indications of [9], a custom questionnaire that was tailored to the user experience with the robot. At the same time, standardised questionnaires were adopted to evaluate the system as a whole, while other custom questionnaires were used to independently evaluate other components of the framework. A full report of the questionnaires can be found in [29]. Here, in accordance with the scope of this paper, we report only the results on the experience with the robot (not discussed in [29]).

6.2 Pilot Results and Quantitative Evaluation

The evaluation we present in this and subsequent sections is based on data extracted from 10 of the 14 total participants of the field study. We decided not to include the data collected from 4 users (2 from the `ITA` group and 2 from the `ESP` one) because their experience was jeopardized by external factors not reflecting the reference scenario sought by our study. Specifically, the reasons were the following:

- a `ESP` user dropped out from the pilot after reporting that the system was not working correctly. Later, we discovered that this was due to poor internet connection (a condition we could not control);
- another `ESP` user experienced similar problems with the internet connection; while this user continued to participate, the robot was unable to move for almost the entire duration of the pilot. Consequently, the data from this user (only a few dozen interventions) were not comparable with those from the others;
- one of the three users from `ITA` living inside the AL facility, despite initially showing interests to the study, did not use the robot and kept the system off for the entire duration. This user was the oldest among the participants (92 years old) and our platform turned out to be not suitable for his needs;
- another `ITA` user living inside the AL facility started the pilot a week before the national COVID-19 lockdown. In order to comply with the imposed restrictions, the management of the AL facility decided to turn off the robot.

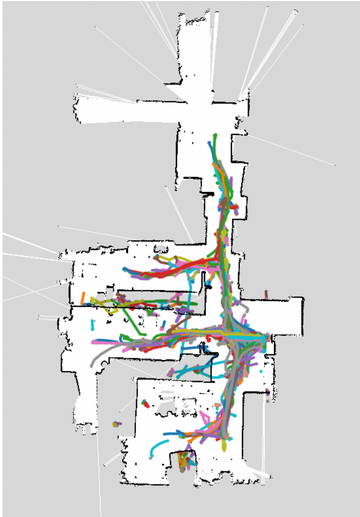


Fig. 9: User `ITA`-1. The paths executed by the robot while moving during the pilot are shown with different colors. This robot lost its localisation 50 times in 1767 m covered.

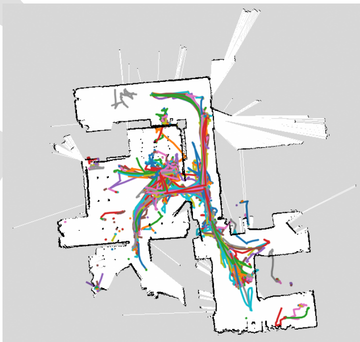


Fig. 10: User `ESP`-3. This robot lost its localisation 79 times in 1337 m covered.

Also, during the 2 pilot rounds, 3 robots had to be replaced due to a hardware failure of the motor control board and battery. The replacement and fix of the robots required approximately 2 weeks each. This time is not computed as part of the pilot time.

Table 3 lists the usage metrics of the 10 robots, showing how they were able to autonomously

Users		ITA-1	ITA-2	ITA-3	ITA-4	ITA-5	ESP-1	ESP-2	ESP-3	ESP-4	ESP-5	ITA	ESP	ALL
Users' Age		76	74	68	81	75	80	84	79	65	71	M=74.8	M=75.8	M=75.3
Pilot Duration (Weeks)		16	16	10	11	11	23	13	12	10	10	64	68	132
Days with an intervention		56	68	39	35	44	86	70	42	45	49	242	292	534
Robot Performances	Distance covered (m)	1767.22	2744.62	1342.88	572.28	1144	1440.39	1800.81	1337.02	687.55	1046.65	7571 (M=5110.77)	6312 (M=1514.2)	13883.42 (M=1388.34)
	Robot moving (Hours)	31	114.98	90.765	9.58	14.55	131.61	28.17	28.94	14.99	23.51	260.875 (M=52.17)	227.22 (M=45.44)	488.095 (M=48.81)
	Lost localisation	50	51	37	47	32	68	36	79	13	42	217 (M=43.4)	238 (M=47.6)	455 (M=45.5)
	Time robot Idle (%)	98.85%	95.72%	94.60%	99.48%	99.21%	96.59%	98.71%	98.56%	99.11%	98.60%	97.57%	98.29%	97.97%
Intervention Performances	N Total	209	208	72	117	116	284	162	141	74	156	722	817	1539
	N Successful	186	178	49	55	80	217	120	90	52	103	548	582	1130
	N Failed	11	23	14	1	14	21	26	23	10	34	63	114	177
	N Not Performed	12	7	9	61	22	46	16	28	12	19	111	121	232
	Performed (%)	89%	85.58%	68.06%	47.01%	68.97%	76.41%	74.074%	63.83%	70.27	66.03%	75.9%	71.24%	73.42%
	Success Rate (%)	94.42%	88.56%	77.78%	98.21%	85.11%	91.18%	82.19%	79.65%	83.87%	75.18%	89.69%	83.62%	86.46%

Table 3: The summary of the statistics concerning the pilot. The Intervention Performances take into consideration the data derived from WM, SQ, and REM interventions as they are the most similar in duration and delivery method. The Robot Performances take into consideration the pilot data as a whole. Whenever it is not specified, the aggregation of the data of the last three columns (ITA, ESP, ALL) is carried out by simple summation. Otherwise, the mean value is reported (M).

move in the challenging context of a long-term unsupervised deployment inside an apartment. The table also reports users' age and the pilot duration, in weeks, for each of the users. On average, each robot travelled 1388.34 m, with a total distance travelled of almost 14 km. During the pilot, the Giraff-X robots were moving inside the apartments for approximately 20 days combined, with an average of 48.8 h per robot, a little more than 2 days for each user. Interestingly, while the robots were active for approximately 1000 days, they performed at least one complete intervention only in 534 days. From one side, this was a result of a specific policy of the VC, which had the goal of not being too invasive and overload the user with many interventions. From the other, interventions were performed, as per users' request, from 9:00 to 19:00. It was often the case that the users were away from their apartments during those hours. As a consequence, the interventions scheduled for that day were postponed to the day after.

Another trend emerging from the stats is that the robots were *idle* at the docking station most of the time, 97.98% on average. This suggests how, towards long-term autonomy, it is important to carefully design a decision-making process that allows the robot to react to unexpected events occurring when *idle*.

Despite the significant amount of time spent by the robots moving inside complex environments, they did not cause any damage to the apartments' furniture, nor they entered in collision with the users. This shows that the process of prepiloting, testing, and refining the navigation abilities of the robot resulted in our system achieving robust navigation performance with respect to the task assigned to the platform.

Fig. 9 and Fig. 10 provide two examples of how navigation performance and robustness are of primary importance in long-term deployments. They show the paths executed by the robots of users ITA-1 and ESP-3. Note that, as we do not have the exact localisation of the robot, we rely on and plot the estimated position made available by the robot itself. Consequently, trajectories plotted in "odd" parts of the map are due to localisation errors that might have occurred. For example, in Fig. 9, the robot sometimes incorrectly estimated that its location was between two rooms, while in Fig. 10 the robot experienced several localisation failures. These latter cases occurred mostly in the central room, connected to the corridor, where its docking station was located. This room, with a dining table and several chairs (whose legs are visible on the map), was rarely used, so the robot rarely found the user there. Nevertheless, the nature of the furniture in that room made robust navigation particularly complex to achieve. It can be seen that, during the approximately 3 km travelled by the two robots combined, they managed to cover almost the entire mapped environment.

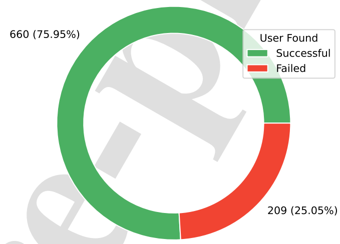


Fig. 11: Performance of the robots in finding the user searching in their apartment.

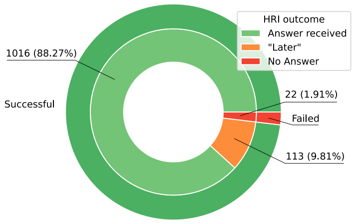


Fig. 12: HRI performance.

6.2.1 Robot Functionalities Evaluation

In this section, we evaluate two core capabilities of the robots: *searching* the user (performed as described in Section 3.1.3) and *interacting* with them (Section 3.1.4).

As already discussed, the task of finding users inside their own homes is particularly difficult. First, it requires robust navigation, integration with IoT sensors (which provide an estimate of the current user location), human detection, and a safe approach to the user. Second, it must often be performed multiple times during a single intervention, since the user might move between the time the identification is completed and the interaction starts. Fig. 11 shows how our SARs were able to correctly search and identify the location of users with a success rate of 75.95%. Note that this rate does not take into account those situations in which the robot immediately finds the user from an intervention completed shortly before (in such cases, the search behaviour is not even initiated). As the exact location of the user is unknown, it is not possible to distinguish the causes that led the robot not to succeed in the search. While the robot may not be able to detect the user due to a failure in one of its components (e.g., losing localisation) or a wrong estimate by the system (e.g., the IoT signals that the user is in some room when actually outdoors), it may also happen that the user could not be identified due to other circumstances (e.g., it is in one of the rooms in which the robot is not allowed to enter). However, despite the inherent difficulty of searching for users who move freely in an apartment, our framework proved to be robust enough to be able to identify them in most cases.

Once the robot has identified and approached the user, the search procedure ends and the HRI module starts. Therefore, in all the occurrences used to evaluate the HRI, the robot is standing in front of the users and have their attention. Fig. 12 shows the performance of the HRI module in receiving an answer from the user after asking a question, asking if they are willing to perform a scenario, or communicating a message. It can be seen how the robot was able to correctly interact with the user in 98.09% of the occasions. In 88.27% of these cases, users decided to interact with the robot, in another 9.81% users responded that they were busy and asked to reschedule the intervention (indicated as "Later" in Fig. 12). Only in less than 2% of the cases the robot tried to interact with the users but it was not able to detect the users' answer, and the interaction failed. Also in this case, the reasons for a failure during an interaction can be different, but the exact cause of a failed HRI is not observable from our data. Examples of such causes are: the robot being unable to detect the user's utterance (e.g., due to noise), the user never answering, or even moving away from the robot after the HRI is started. However, from our data we can see that, despite finding users is challenging, once they are found and engaged by the robot the (vocal) interaction is very likely to succeed.

6.2.2 Robot Intervention Evaluation

In this section, we evaluate the performance of the SARs in executing the interventions. To do so, we consider three types of intervention: Spot Question (SQ), Weight Measurement (WM), and Reminder (REM), as described in Table 2. We decided to take into consideration these interventions because they are the most performed ones by the robot, they all require the robot to find the user and interact with them, they have comparable duration, and they are triggered by the VC.

We consider an intervention as *successful* when the robot concludes its interaction with the user with profit, otherwise we consider the intervention as *failed*. Table 3 shows the total number of interventions performed and their outcome. We present average results obtained across all users and detailed results obtained for each user. At the same time, Fig. 13 presents the results of the interventions (on average) and the main reasons for the different outcomes.

It can be seen how the robot tried to complete, on average, 72.42% of the total interventions requested by the system. Interventions that are not performed, either not executed at all by the robot or cancelled during their execution by an external event, are rescheduled for the next days by the VC. We consider the fact that some interventions were not performed as a positive outcome, as the robot

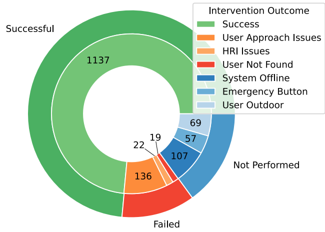


Fig. 13: Performance of the robot in completing an interventions, divided per outcome reason.

was able to detect that the circumstances did not allow the successful completion of the requested intervention. As an example, approximately 7% of the interventions were not performed due to the fact that the system was offline. This was caused either by the user disabling the system or by internet connectivity issues. Similarly, 4.56% of interventions were not executed due to the fact that the users were signalled as *outdoor* by the IoT system. Note that, if an intervention is cancelled and set to "not performed" after the robot is already moving, we resorted to the safe behaviour of sending the robot back to docking station.

Overall, robots performance was satisfactory, as they were able to carry out the assigned intervention successfully in 86.46% of the cases (Success Rate in Table 3).

The causes that led to failures during the execution of the interventions were different in nature. In approximately 10% of the cases, the robot failed to approach the user after identifying the users location because either they changed their location (e.g., by exiting the robot's field of view), due to a navigation failure (e.g., the robot went too close to an obstacle), or due to a localisation failure. In 1.5% of cases, the robot was never able to find the user, and in 1.5% of cases the robot was not able to understand the answer of the user or to talk to them.

6.3 On-the-field Case Studies

Despite the efforts to devise a robust platform, the particularly challenging setting of the pilot resulted in a number of critical and unexpected contingencies. Upon such events, users largely exploited the possibility of reaching for remote support, especially in the early days of the pilot deployment, to report problems and receive explanations about what might have caused robot's anomalous behaviours. The MLS was the central component that made this assistance activity possible. In this section, we report some examples that provide informative insights on the kind of difficulties that could be encountered during a long-term deployment.

Critical Night Failure At the beginning of the first round, the user `ITA-5` experienced a failure in the IoT network: one of the microphones wrongly detected a "call for help" request at midnight. In response to this wrongly detected need, the robot tried to locate the user to confirm the request. The unsuccessful search lasted for approximately half an hour (as the robot was unable to reach the user who was in the bedroom with the door closed); eventually, the robot was unable to return to the docking station due to being stuck inside a room by a half-closed door, asking by voice for external support. The user was awakened by the robot's request. He located Giraff-X, cancelled the intervention, and went back to bed, leaving the robot unable to move. After 2 hours, the robot, still unable to move where the user left it, asked again the user for help to reach its docking station as the battery level was critical. The user was awakened again by the robot's voice and brought back the robot to the docking station. The following morning, `ITA-5` contacted our technical team. The MLS allowed us to replay the sequence of events that happened during the night, eventually revealing what happened. The cause of the failure was explained to the user, and he was reassured. In the next days, the technical team was able to provide additional fixes to prevent such behaviours from happening again. As a result, the user accepted the explanations and decided to continue with the pilot. Other users signalled during the initial phased of the pilot similar, yet less intrusive, events (e.g., the robot talking during the night due to a power failure). Following this, a more robust methodology to prevent the robot to perform unwanted interventions at night has been implemented.

Robot Requests for Support A few days later, `ITA-5` contacted the technical team to report that the robot was standing motionless in the middle of the room after performing an intervention. `ITA-5` reported that he was neither confident nor safe after the previous incident. By inspecting the MLS, we discovered that the robot was working correctly and that the issue was that `ITA-5` was overseeing the robot's movements by standing right in the middle of the path the robot should have followed towards the docking station, blocking its trajectory and forcing it to stay motionless in that point. We instructed the user to clear the robot's path, allowing the robot to recover its motion. The robot successfully returned to the docking station and the user was reassured about the proper functioning of the system. After this, we received several comments from users who signaled that, after talking to them, the robot often remained static in their position for some minutes. We then explained that it was because they were blocking their path towards the docking station. As a consequence of their better understanding of the system, the users adapted their behaviours to avoid the repetition of the problem, and they were not worried by such events anymore.

Robot Moving while User Away `ITA`-5 contacted us to signal that the robot was wrongfully moving around his house while he was not there and that, for this reason, he was not feeling safe leaving his house. We inspected the MLS, through its replay functionality, and we observed that while the user was away, the robot was behaving correctly and not moving. `ITA`-5 was referring to a one-time situation in which the user, who had been away for the entire day, re-entered the apartment just to immediately leave again for a couple of minutes. The short presence triggered a robot intervention that was then cancelled since the user was detected to no longer be at home. Upon re-entrance a few minutes later, the user saw the robot moving and suspected some kind of activity taking place when he was out, not realising that it was his previous entrance that triggered it. After the explanation, `ITA`-5 understood what happened, and was confident about the proper functioning of the robot.

Similar behaviour was signaled by `ITA`-1. Upon her return home after a few days away, she signalled that she found the robot switched off in the middle of a room. By inspecting the MLS through its replay functionality, we discovered how the robot lost connection to its docking station due to an external event and tried to trigger a docking manoeuvre. However, as the house was in total darkness, the docking approach failed, and the robot shut down to preserve its batteries. The user was convinced by the explanation and was reassured that the robot was not moving without her approval or in her absence. Overall, several users signalled events where the robot was moving while they were away. Proving detailed explanation by exploiting the MLS was key in reassuring the users and preserve their trust.

HRI Inconsistencies in Interventions Two of the interventions carried out by the robot were particularly problematic from the HRI point of view. This was not due to errors in the HRI module, but rather in the requests the robot was making to the users per se. The first iteration of the WM intervention was expected to ask the users to weight themselves on the smart scale before breakfast and wearing only their pyjamas. Even though the VC scheduled the intervention to be performed as the first thing in the morning, it was not possible to verify that the users hadn't already had breakfast nor that they were still wearing their nightwear. This led to multiple instances where the users contacted the technical team to complain about the request of the robot and their impossibility to comply, which was considered very frustrating. After collecting these comments from the users, the intervention was updated so that the robot would just ask them to weight themselves, without additional details as to when or how.

Similarly, one of the type of Spot Question (SQ) considered important by clinicians was that of *confabulation*. According to this type of SQ, the robot had to find the users and ask them random open-ended questions (e.g., "Do you remember what was wearing the person seated next to you last time you took the bus?"). The aim was to let the users speak and record their confabulation on the topic of the question. However, the randomness of the questions was not appreciated by the users, who did not understand what and why the robot would ask them. For this reason, they often contacted the technical team to report that the robot "had gone insane" or was broken. Only by analysing what happened through the MLS and double-checking with the users, it was possible to understand that the problem was that particular kind of SQ intervention, which was immediately dropped to avoid frustrating the users even further.

Remote Monitoring During COVID-19 Lockdown Most of the second round of the pilot took place during the COVID-19 national lockdown, preventing any form of in-place support. During that period, users who lived alone spent all day at home with their robots. Users were asked if they wanted to continue with the pilot even in this particular situation, and all agreed. Thanks to the constant feedback they gave to technicians and thanks to the use of the MLS, we were able to provide support and allow the users to successfully complete the pilot.

Moreover, user ITA-2 explicitly requested to have the system on and remotely maintained even after the official pilot period. This user was particularly engaged with the robot and was happy to "see the robot wandering around the house". Overall, the user kept the system and the robot active for two additional months after the end of the pilot, providing positive feedback and reporting that the system and the presence of the robot, "was of great support during the difficult time".

6.4 Robot Acceptance Questionnaire

At the end of the evaluation period, our users were asked to answer a set of questionnaires related to their experience (using a 1 to 5 Likert scale, with 1 for "Strongly Disagree" and 5 for "Strongly Agree"). The questionnaires were developed, along with the clinical partners of the project, to evaluate each component of MoveCare and their acceptability with the goal of immediately identifying improvement actions for each of them. Additional standardised questionnaires were adopted to evaluate the system as a whole and are reported in [29]. Table 4 shows the results gathered through the questions pertinent to the robot. A full report of the questionnaire answers, provided for completion and clarification purposes, divided by pilot country is in Appendix A, Table A1.

The proposed framework, with the use of the MLS to provide remote support, allowed participants to report feeling safe and comfortable both during robot navigation and during human-robot interaction, as can be seen from the answers provided to questions Q[1-5]. Despite the size of the robot, the users did not feel that it was intrusive (Q5) or that it was a danger to their apartment (Q4). This suggests that a more conservative behaviour ensuring safe navigation is appreciated by the end-users. From the answers in Table 4, it can be seen how the robot was accepted by the users, who understood and positively evaluated the robot behaviours. As an example, User ITA-5, even after the critical events reported in the previous section, answered with the highest mark to questions Q[1-5]. Despite these positive remarks, users reported that the robot should be more helpful and responsive, highlighting that there is still work to do to meet the needs of the potential end-users of SAR-based systems.

In particular, by comparing the results of the quantitative data collected during the pilot and the data of the questionnaire, it is easy to see that there are some discrepancies. Indeed, the questions related to the navigation and approach capabilities of the robot are consistently rated very high by users (Q1, Q2, Q4), even though they were not always performing correctly. We can see that, even if navigation sometimes failed or took a long time to be completed, users always felt safe and comfortable around the robot while it was navigating. This was reflected in a high evaluation of the robot as a whole, as the perceived safety of the robot navigation made sure that the users would not worry about the robot moving, thus accepting their presence. On the other hand, while they also felt comfortable when interacting with the robot and the HRI module performed better in terms of failed interactions, delays in communication and slow responsiveness detected by pilot users made it so that the robot interface received generally lower scores (Q10, Q13, Q15), with the lowest scores given by the oldest participants. This underlines how important the trade-off between performance and user requirements is, and how valuable the work of integrating the users' feedback in the development of SAR is. It also suggests that users are generally less lenient towards the HRI module of the robot, which is more susceptible to minor issues with respect to other core functionalities of SARs, as also underlined by the case studies of the previous section.

Furthermore, we have investigated (using Mann-Whitney U Test) whether the country of the pilot affected the results of the questionnaire, without finding any statistical significance when analysing the answers by country. This could be due to the very small sample (5 Italian users and 5 Spanish users). Looking at the answers in Table 4 it could be seen that the answers to Q6 are the ones that change drastically by country. This is probably due to the fact that that particular question is somewhat ambiguous, especially in its translation in Italian and Spanish, a fact that we noticed only ex-post.

Overall, our system was positively evaluated and its usefulness was understood by prospective users. To further sustain the need for the robot in such contexts the reader is referred to [29], where we show how the robot had a major impact in the total usage of the system.

	Question	M ITA	M ESP	M
Q1	While the robot is navigating, I am feeling safe	5	5	5
Q2	While the robot is navigating, I am feeling comfortable	5	5	5
Q3	The robot is not intrusive during the environmental exploration	4	3	3.5
Q4	While the robot is navigating, I am not feeling in apprehension for the objects in the house	5	5	5
Q5	I don't feel pressured when the robot approaches me	5	5	5
Q6	Autonomous robot navigation helps me in everyday life tasks	1	4	2
Q7	I clearly understand when the robot is approaching me	4	4	4
Q8	The robot detects me quickly	2	4	2.5
Q9	The presence of the robot helps me in managing emergency situations with a lower level of anxiety	3	4	3
Q10	Vocal interface is responsive to my inputs	2	4	3
Q11	Vocal instructions provided by the robot are clear	4	5	4.5
Q12	Vocal reminders provided by the robot are clear	4	4	4
Q13	I can easily interact with the robot through vocal commands	2	3	2.5
Q14	I am feeling comfortable when interacting with the robot	3	4	4
Q15	When interacting with the robot, I don't perceive delays between its requests and its answers	2	2	2
Q16	I think that the feedback from the robot should be more informative	4	3	3.5

Table 4: Robot Acceptance Questionnaire. M is the Median, divided per country in ITA and ESP.

7 Lessons Learned and Discussion

The experience obtained during our long-term experimental campaign, supported by our MLS tool, allowed us to improve several features of the system. Table A2, in Appendix, provides notable examples of such improvements, together with their rationale, in the attempt of providing useful hints for future practitioners in the field. In this section, we try to generalize from those examples, and the results presented above, by providing a discussion around the lessons we learned and open questions.

7.1 Pilot Organisation and Testing

Preliminary experimental campaigns are a precondition to a robust system. The prepilot testing described in Section 5 allowed us to refine several components of the system into more robust solutions, revealing unexpected contingencies and taking into account the feedback from users. Nonetheless, critical situations proper of a long-term deployment had a very low probability of occurring during these tests. For example, the night failure described in Section 6.3, was due to a combination of factors: a false detection of a call for help, an unreachable user, and the night time. An appropriate system response had been designed and tested for each of these factors individually. However, the occurrence of all of them at the same time is very unlikely in the short term, and it did not happen during any of the prepilots. The likelihood of such events becomes non-negligible as the deployed systems in the home of end-users become increasingly complex, integrated, and employed for longer periods of time. The availability of well-defined methods, perhaps supported by benchmarks, to deal with such complex combinations of events is central to acceptability but still an open problem for mobile SARs research.

The MLS has played a primary role in closing the loop between the users and supervisors of the system. It allowed us to remotely intervene on specific issues (a crucial capability at the beginning of the COVID-19 pandemic, where users' houses could not be visited) and understand unexpected events (such as those described in Section 6.3). Ultimately, this gave us additional knowledge on the challenges SARs must face in the long-term use.

One additional feature of the MLS is its possibility of being accessible not only to system designers but also to external non-expert operators. The campaign we carried out in Italy was managed by researchers who contributed to the system design and, therefore, had a deep understanding of it. Instead, the Spanish pilot was managed by external contractors who did not participate in the development of the system and were only briefed about it in a couple of sessions. These contractors, especially during the initial phases of the study, relied heavily on the MLS not only to supervise the system, but also to learn about its functionalities and limitations. At the end of the study, they quoted the MLS as an essential tool.

Finally, it is also important to highlight the intrinsic ethical challenges that a data harvesting process like the one operated by the MLS raises in the domain of assistive technologies for welfare [40]. We did not perform any additional reasoning or pattern analysis, but only provided an interface for logged data.

7.2 The Technical Side

Most of the effort spent developing a SAR goes into functionalities for task execution such as navigation or user detection. However, our study brought to light an obvious but underestimated issue, namely that the vast majority of operational time is *not* spent performing any of such tasks. Instead, the robot spends almost all of its time in an idle status, waiting (see Table 3). This is a consequence of the long-term setting, where interventions naturally assume a profile characterized by some temporal sparsity, in part due to the combination of the system's long persistence and its requirement of not being too invasive. While it is true that the robot is not supposed to perform any task (besides charging) while idle, the extensive duration of the idle time increases the chances of unexpected events to which the robot may have to respond, restoring its operative condition or exiting the idle state (e.g., in case of a call for help). For example, the robot might disconnect from the docking station, or a power or internet outage might occur. These events are challenging not only to foresee, but also to model and detect. Even though the robot should ideally be able to autonomously recover from such events, in practise it is often impossible. Interestingly, we noted that most of the critical problems reported by the users were related precisely to such unexpected events. These findings highlight the importance of not only developing the intended tasks of the system but also explicitly modelling for unexpected events and carefully devising robust recovery behaviours.

An example of an unexpected yet particularly important critical issue for our platform was the lack of stable internet connectivity. While in preliminary testing this feature was taken for granted, we quickly discovered how encountering issues with it significantly jeopardised the robot's performances. The vocal interactions with the users were particularly affected by connectivity problems, as they introduced a substantial delay in speech recognition that caused the interactions to be characterised by long pauses from the robot. This problem was particularly perceived by the users, as reflected in the answers to the questionnaire in Table 4, Q12 and Q15. This issue was particularly critical (leading to the only drop-out from the study) as it was not under our control and, consequently, not fixable. As a result, we introduced several countermeasures to cope with this issue, without being able to fully solve it.

7.3 Explainability and Acceptability

Much effort has been spent in making the interactions between the user and the robot during interventions as much intuitive and easy as possible. However, we realised that interactions were also taking place when the robot was in an idle state or in a self-management task. A frequent example was the robot remaining immobile for several minutes since the path to its docking station was blocked (by the user, by an object, ...). In these cases, users often believed that the robot was not working. Adding explainability to these behaviours, while ensuring that the robot did not become too intrusive, is a desirable feature for which no clear solution is available. One possible approach is to let the robot itself notify the user of its failures, for instance when it lost its localisation.

We also learned an important lesson about acceptability. While a good performance in all the robot's abilities is essential, these abilities are not viewed equally by the users. Examples are navigation and human-robot interaction abilities, as explained in Section 6.4. The user's perception must be taken into account when interpreting other quantitative performance metrics. Consider, for example, robot navigation. Our study highlighted how users cared about not feeling threatened by the robot's movement. Our navigation module, whose initial performance was satisfactory, did not account for such a factor despite having a remarkable impact on it. Our proposed solution was to improve the navigation module to obtain conservative paths, followed slowly, and without hazardous movements. Instead of taking risks, the robot preferred choice was to try to find safer paths, eventually resorting to recovery behaviour or stop moving, asking the user for help.

As a result, robot movements were perceived as more predictable and the platforms were able to not cause any damage in the users' apartments. As discussed in Section 6.4, this was appreciated by users. However, such a conservative policy increased the chances of the robot being stuck during navigation, a situation that could have been avoided had users' perception not been taken into account. Nevertheless, this seems not to have negatively impacted robot acceptability.

7.4 Long-term HRI

Having a robot "living" with users for several weeks allowed us to reason about the impact of SARs human-robot interaction in the long term. The positive remarks made by our users in the questionnaire and their feedback during the campaign suggest how SARs could be positively perceived even after the initial novelty effect. One key factor that contributed to this was the robot's overall discreet presence: the number of performed interventions per day had always been kept low and their executions took place only in working hours.

This finding is somehow in contrast with recent trends, where SARs are presented as persistent companions that engage frequent and sustained interactions. Our observations, instead, indicate how adaptability trumps the enforcement of schedules. A robot that can adapt its behaviour to the daily living habits of its users by performing a few well-targeted interventions per day can be better accepted. This is supported also by the common request made by the users to have the possibility of explicitly request the services of the robots (e.g., asking the robot to reach them, or to go somewhere) alongside the interventions triggered by the system.

7.5 Robot Performance

On several occasions, the robot lost its localisation and undertook recovery behaviours to restore it. Most of the time, however, this was not sufficient, and the robot was forced to ask the assistance of its user. Understanding the causes of these events is of paramount importance.

A first cause can be ascribed to the users: some of their actions could increase the difficulty for the robot to complete a task. The most typical example was that of the user blocking the path of the robot. Another reason that made localisation struggle was the presence of challenging areas in the apartments. Specifically, these were regions with dynamic features (objects or furniture that was frequently moved) or parts of the maps that were very similar to others. Fig. 14 shows an example of this. On the right, it can be observed where the robot wrongly localised itself repeated times in a challenging area around table and chairs. Although the location is clearly identifiable, the actual specific reasons for the lost localisation are not ascertained. Identify and modelling them could significantly increase robustness and stability, and represents an intriguing future direction for research. As we have already pointed out in Section 7.3, one of the central challenges is the assessment and even the definition of failures,

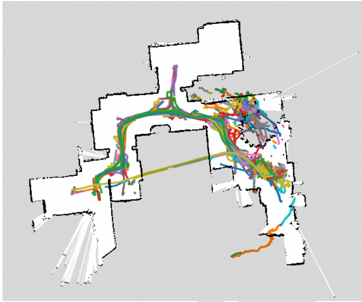


Fig. 14: User `ESP`-1. This robot lost its localisation 68 times in 1440 m covered.

successes, and robust performance, where the process of taking into account the user's perception cannot be neglected.

7.6 Learning from Failures

During the deployment of the system, we observed how most of the interventions were not only successful but also quick: the robot was able to quickly find the user and approach them reliably using a combination of environmental sensors and computer vision. Failed interventions were usually much longer, with the robot wandering across different rooms in search for the user, a search that ultimately did not succeed. It is impossible to determine whether these failures were due to an error in the user's localisation from the environmental sensors, a failure in detecting the user through the robot's camera, or a combination of both. This observation opens different paths for future development. First, since finding the user reliably is of paramount importance, particular care must be taken with methods and algorithms enabling robust user detection. This is even more important as an incorrect detection leading to a failed intervention turned out to be particularly costly since the robot travelled abundantly, thus increasing the chances of getting stuck, running out of battery, or experiencing other motion-related failures. The data depicted in Fig. 15 support this consideration. In Fig. 15, we show the number of times the robot was able to detect a user, the number of times it tried to approach them, the number of rooms traversed by the robot while searching for a user, and the time it spent navigating for both successful and failed interventions. It can be seen that in failed interventions, the robots almost never de-

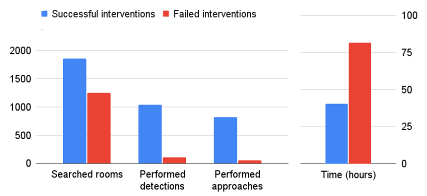


Fig. 15: Robot activities during successful and failed interventions.

tected or approached the users but spent a lot of time navigating.

In general, if the intervention has to fail, it is better for it to fail quickly. Recovery behaviours are the key to this. In our system, if the robot could not find the user in the room indicated by the user-location module, its recovery behaviour would be to roam all the other rooms in search for them. Our deployment has, however, shown that this is unlikely to succeed. One improvement could hence be for the robot to drop the intervention after a predetermined amount of time, letting the VC reschedule the intervention later. Moreover, improved versions of the system could also maintain estimates on the typical success rate of a given intervention through time. In this way, the robot can properly evaluate, in an online fashion, the chances of success and the cost of multiple repeated attempts and act accordingly.

8 Conclusions

In this work, we presented the results obtained in a long-term experimental campaign during which a fleet of SARs was used to support and assist older adults living independently and alone. We showed how the effort put into the different steps needed to develop such complex platforms led to robust systems that performed adequately in both labs and apartments for older adults. Importantly, we show how prepilots were fundamental to begin integrating user feedback into the development of SARs for a longer and successful campaign. We also showed how a system that provides remote supervision of an autonomous SAR-based system increased its level of control and explainability and, ultimately, improved the acceptance and confidence of the end-users. The emerging feedback loop between users, technicians, and caregivers initiated by the Monitoring and Logging System revealed fundamental for the success of the experimental campaign, both in its function of reassuring the users and diagnosing newly discovered issues, which could therefore be fixed. Additionally, we leave to the community a series of lessons learnt that we hope could act as guidelines for the future advancements of SARs. Future work will investigate the possibility of performing online anomaly detection (e.g., as we did in the preliminary work of [2] using prepilot data) and resolution of critical events and learning from robot experience to improve their long-term autonomy, as discussed in Section 7.6. Despite the problems encountered during the pilot, the campaign was a success and gave us great insights on the future development of mobile assistive robots. This work represents an additional step leading towards the widespread adoption of SARs for independent living.

Data and Code Availability Statement

Further details upon the methods used for the collection of these data and their content are available upon request made to the corresponding author. All data are available to the corresponding author for further inquiries. The anonymised version of data that support the findings of this study are available on request from the corresponding author, ML. The data are not publicly available due to restrictions, as they contain information that could compromise the privacy of research participants. The code is available upon requests.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

The research leading to these results received funding from European Commission H2020 - project MoveCare - under Grant Agreement No ICT-26-2016b – GA 732158 and project Essence grant SC1-PHE-CORONAVIRUS-2020-2B - GA 101016112, and the Italian PON project SI-Robotics.

Compliance with Ethical Standards:

The pilot has been approved by the ethical committees of Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy (8/04/19) and of Gerencia del Area de Salud de Badajoz, Junta de Extremadura, Badajoz, Spain (13/02/19).

References

1. Movecare: Multiple-actors virtual empathic caregiver for the elder. URL http://www.movecare-project.eu/
2. Azzalini, D., Castellini, A., Luperto, M., Farinelli, A., Amigoni, F.: HMMs for anomaly detection in autonomous robots. In: International Conference on Autonomous Agents and MultiAgent Systems AAMAS, pp. 105–113. ACM (2020)
3. Bajones, M., Fischinger, D., Weiss, A., Others: Results of field trials with a mobile service robot for older adults in 16 private households. J. Hum.-Robot Interact. **9**(2) (2019)
4. Biswas, J., Veloso, M.: The 1,000-km challenge: Insights and quantitative and qualitative results. IEEE Intelligent Systems **31**(3), 86–96 (2016)
5. Cook, D.J.: Health monitoring and assistance to support aging in place. Universal Computer Science **12**(1), 15–29 (2006)
6. Coradeschi, S., Cesta, A., Cortellessa, G., Coraci, L., Galindo, C., González-Jiménez, J., Karlsson, L., Forsberg, A., Frennert, S., Furfari, F., Loutfi, A., Orlandini, A., Palumbo, F., Pecora, F., Von Rump, S., Štimec, A., Ullberg, J., Ötslund, B.: Giraffplus: A system for monitoring activities and physiological parameters and promoting social interaction for elderly. Advances in Intelligent Systems and Computing **300**, 261–271 (2014)
7. Cortellessa, G., De Benedictis, R., Fracasso, F., Orlandini, A., Umbrico, A., Cesta, A.: Ai and robotics to help older adults: Revisiting projects in search of lessons learned. Paladyn, Journal of Behavioral Robotics **12**(1), 356–378 (2021)
8. Coşar, S., Fernandez-Carmona, M., Agrigoroaie, R., et al.: Enrichme: Perception and interaction of an assistive robot for the elderly at home. International Journal of Social Robotics **12**(3), 779–805 (2020)
9. Daniele, K., Marcucci, M., Cattaneo, C., Borghese, N.A., Zannini, L.: How prefrail older people living alone perceive information and communications technology and what they would ask a robot for: Qualitative study. Journal of medical Internet research **21**(8), e13228 (2019)
10. Feil-Seifer, D., Mataric, M.J.: Socially assistive robotics. IEEE Robotics and Automation Magazine **18**(1), 24–31 (2011)
11. Fiorini, L., Esposito, R., Bonaccorsi, M., Petrazzuolo, C., Saponara, F., Giannantonio, R., Petris, G.D., Dario, P., Cavallo, F.: Enabling personalised medical support for chronic disease management through a hybrid robot-cloud approach. Autonomous Robots **41**(5), 1263–1276 (2017)
12. Fiorini, L., Sorrentino, A., Becchimanzi, C., Mattia, P., Tosi, F., Cavallo, F.: Living with a telepresence robot: results from a field-trial. IEEE Robotics and Automation Letters (2022)
13. Fischinger, D., Einramhof, P., Papoutsakis, K., Wohlkinger, W., Mayer, P., Panek, P., Hofmann, S., Koertner, T., Weiss, A., Argyros, A., Vincze, M.: Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. Robotics and Autonomous Systems **75**, 60–78 (2016)
14. Garcia-Salguero, M., Monroy, J., Solano, A., Gonzalez-Jimenez, J.: Socially acceptable approach to humans by a mobile robot. In: International Conference on Applications of Intelligent Systems (APPIS), pp. 1–7 (2019)
15. García-Soler, Á., Facal, D., Díaz-Orueta, U., Pigini, L., Blasi, L., Qiu, R.: Inclusion of service robots in the daily lives of frail older users: a step-by-step definition procedure on users' requirements. Archives of Gerontology and Geriatrics **74**, 191–196 (2018)
16. Gonzalez-Jimenez, J., Galindo, C., Ruiz-Sarmiento, J.R.: Technical improvements of the giraff telepresence robot based on users' evaluation. In: IEEE International Symposium on Robot and Human Interactive Communication, pp. 827–832 (2012)
17. Gross, H.M., Mueller, S., Schroeter, C., Volkhardt, M., Scheidig, A., Debes, K., Richter, K., Doering, N.: Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5992–5999 (2015)
18. Gross, H.M., Scheidig, A., Müller, S., Schütz, B., Fricke, C., Meyer, S.: Living with a mobile companion robot in your own apartment-final implementation and results of a 20-weeks field study with 20 seniors. In: International Conference on Robotics and Automation (ICRA), pp. 2253–2259 (2019)
19. Gross, H.M., Schroeter, C., Mueller, S., Volkhardt, M., Einhorn, E., Bley, A., Langner, T., Merten, M., Huijnen, C., Heuvel, H.v.d., Berlo, A.v.: Further progress towards a home robot companion for people with mild cognitive impairment. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 637–644 (2012)
20. Hawes, N., Burbridge, C., Jovan, F., Kunze, L., Lacerda, B., Mudrova, L., et al.: The strands project: Long-term autonomy in everyday environments. IEEE Robotics Automation Magazine **24**(3), 146–156 (2017)
21. Kunze, L., Hawes, N., Duckett, T., Hanheide, M., Krajník, T.: Artificial intelligence for long-term robot autonomy: A survey. IEEE Robotics and Automation Letters **3**(4), 4023–4030 (2018)
22. Lammer, L., Huber, A., Weiss, A., Vincze, M.: Mutual care: How older adults react when they should help their care robot. In: Proc NF-HRI (2014)
23. Lunardini, F., Borghese, N.A., Piccini, L., Bernardelli, G., Cesari, M., Ferrante, S.: Validity and usability of a smart ball–driven serious game to monitor grip strength in independent elderlies. Health Informatics Journal **26**(3), 1952–1968 (2020)
24. Lunardini, F., Febbo, D.D., Malavolti, M., Cid, M., Serra, M., Piccini, L., Pedrocchi, A.L.G., Borghese, N.A., Ferrante, S.: A smart ink pen for the ecological assessment of age-related changes in writing and tremor features. IEEE Transactions on Instrumentation and Measurement **70**, 1–13 (2020)
25. Lunardini, F., Luperto, M., Daniele, K., Basilico, N., Damanti, S., Abbate, C., Mari, D., Cesari, M., Ferrante, S., Borghese, N.A.: Validity of digital trail making test and bells test in elderlies. In: IEEE International Conference on Biomedical Health Informatics (BHI), pp. 1–4 (2019)
26. Lunardini, F., Luperto, M., Romeo, M., Renoux, J., Basilico, N., Krpič, A., Borghese, N.A., Ferrante, S.: The movecare project: Home-based monitoring of frailty. In: IEEE International Conference on Biomedical Health Informatics (BHI) (2019)
27. Lunardini, F., Luperto, M., Romeo, M., et al.: Supervised digital neuropsychological tests for cognitive

decline in older adults: Usability and clinical validity study. JMIR mHealth and uHealth **8**(9), e17963 (2020)

28. Luperto, M., Basilico, N., et al.: A community based activity center to promote social engagement and counteract decline of elders living independently. Lect Notes Artif Int (2021)
29. Luperto, M., Monroy, J., Renoux, J., Lunardini, F., Basilico, N., Bulgheroni, M., Cangelosi, A., Cesari, M., Cid, M., Ianes, A., Gonzalez-Jimenez, J., Kounoudes, A., Mari, D., Prisacariu, V., Savanovic, A., Ferrante, S., Borghese, N.A.: Integrating social assistive robots, iot, virtual communities and smart objects to assist at-home independently living elders: the movecare project. International Journal of Social Robotics (2022). Accepted
30. Luperto, M., Monroy, J., Ruiz-Sarmiento, J.R., Moreno, F.A., Basilico, N., Gonzalez-Jimenez, J., Borghese, N.A.: Towards long-term deployment of a mobile robot for at-home ambient assisted living of the elderly. In: European Conference on Mobile Robots (ECMR), pp. 1–6 (2019)
31. Luperto, M., Romeo, M., Lunardini, F., Basilico, N., Abbate, C., Jones, R., Cangelosi, A., Ferrante, S., Borghese, N.A.: Evaluating the acceptability of assistive robots for early detection of mild cognitive impairment. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1257–1264 (2019)
32. Luperto, M., Romeo, M., Monroy, J., Vuono, A., Basilico, N., Gonzalez-Jimenez, J., Borghese, N.A.: What is my robot doing? remote supervision to support robots for older adults independent living: a field study. In: 2021 European Conference on Mobile Robots (ECMR), pp. 1–7 (2021). doi:10.1109/ECMR50962.2021.9568839
33. Marder-Eppstein, E., Berger, E., Foote, T., Gerkey, B., Konolige, K.: The office marathon: Robust navigation in an indoor office environment. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 300–307 (2010)
34. Moreno, F.A., Monroy, J., Ruiz-Sarmiento, J.R., Galindo, C., Gonzalez-Jimenez, J.: Automatic waypoint generation to improve robot navigation through narrow spaces. Sensors **20**(1), 240 (2020)
35. Orlandini, A., Kristoffersson, A., Almquist, L., Björkman, P., Cesta, A., Cortellessa, G., Galindo, C., Gonzalez-Jimenez, J., Gustafsson, K., Kiselev, A., Loutfi, A., Melendez, F., Nilsson, M., Hedman, L.O., Odontidou, E., Ruiz-Sarmiento, J.R., Scherlund, M., Tiberio, L., von Rump, S., Coradeschi, S.: ExCITE project: A review of forty-two months of robotic telepresence technology evolution. Presence: Teleoper. Virtual Environ. **25**(3), 204–221 (2017)
36. Papadopoulos, C., Castro, N., Nigath, A., et al.: The caresses randomised controlled trial: exploring the health-related impact of culturally competent artificial intelligence embedded into socially assistive robots and tested in older adult care homes. International Journal of Social Robotics **14**(1), 245–256 (2022)
37. Portugal, D., Alvito, P., Christodoulou, E., Samaras, G., Dias, J.: A study on the deployment of a service robot in an elderly care center. International Journal of Social Robotics **11**(2), 317–341 (2019)
38. Renoux, J., Luperto, M., Basilico, N., Romeo, M., Milis, M., Lunardini, F., Ferrante, S., Loutfi, A., Borghese, A.: A virtual caregiver for assisted daily living of pre-frail users. In: German Conference on Artificial Intelligence (KI20), pp. 176–189 (2020)
39. Salatino, C., Gower, V., et al.: The EnrichMe project: A robotic solution for independence and active aging of elderly people with MCI. In: Proc. ICCHP, pp. 326–334 (2016)
40. Sánchez, V.G., Taylor, I., Bing-Jonsson, P.C.: Ethics of smart house welfare technology for older adults: A systematic literature review. International journal of technology assessment in health care **33**(6), 691–699 (2017)
41. Schroeter, C., Mueller, S., Volkhardt, M., Einhorn, E., Huijnen, C., van den Heuvel, H., van Berlo, A., Bley, A., Gross, H.M.: Realization and user evaluation of a companion robot for people with mild cognitive impairments. In: 2013 IEEE International Conference on robotics and automation, pp. 1153–1159. IEEE (2013)
42. Scopelliti, M., Giuliani, M.V., Fornara, F.: Robots in a domestic setting: a psychological approach. Universal Access Inf. Soc. **4**(2), 146–155 (2005)
43. Vuono, A., Luperto, M., Banfi, J., Basilico, N., Borghese, N.A., Sioutis, M., Renoux, J., Loufti, A.: Seeking prevention of cognitive decline in elders via activity suggestion by a virtual caregiver. In: International Conference on Autonomous Agents and MultiAgent Systems, p. 1835–1837 (2018)

A Appendix

We report here the full Robot Acceptance questionnaire, with the answers collected from the 10 pilot participants (5 Italian and 5 Spanish).

The question legend is the following:

- Q1: While the robot is navigating, I am feeling safe
- Q2: While the robot is navigating, I am feeling comfortable
- Q3: The robot is not intrusive during the environmental exploration
- Q4: While the robot is navigating, I am not feeling in apprehension for the objects in the house
- Q5: I don't feel pressured when the robot approaches me
- Q6: Autonomous robot navigation helps me in everyday life tasks
- Q7: I clearly understand when the robot is approaching me
- Q8: The robot detects me quickly
- Q9: The presence of the robot helps me in managing emergency situations with a lower level of anxiety
- Q10: Vocal interface is responsive to my inputs
- Q11: Vocal instructions provided by the robot are clear
- Q12: Vocal reminders provided by the robot are clear
- Q13: I can easily interact with the robot through vocal commands
- Q14: I am feeling comfortable when interacting with the robot
- Q15: When interacting with the robot, I don't perceive delays between its requests and its answers

Question	ITA-1	ITA-2	ITA-3	ITA-4	ITA-5	ESP-1	ESP-2	ESP-3	ESP-4	ESP-5	ITA	ESP	ALL
Q1	5	3	5	2	5	4	5	5	3	5	5	5	5
Q2	5	2	5	3	5	4	5	5	4	5	5	5	5
Q3	5	2	5	2	4	3	4	3	4	3	4	3	3.5
Q4	5	2	5	2	5	5	5	5	4	5	5	5	5
Q5	5	3	5	5	5	4	5	5	4	5	5	5	5
Q6	1	3	1	1	1	1	4	4	4	4	1	4	2
Q7	5	2	2	4	4	3	5	4	4	5	4	4	4
Q8	5	3	2	1	1	4	2	1	4	4	2	4	2.5
Q9	2	3	3	1	5	1	3	4	4	4	3	4	3
Q10	5	3	2	1	2	4	4	3	4	2	2	4	3
Q11	5	3	5	3	4	4	5	5	4	5	4	5	4.5
Q12	5	2	5	3	4	4	1	5	4	5	4	4	4
Q13	5	2	1	1	2	4	1	3	3	3	2	3	2.5
Q14	5	2	3	3	4	4	2	5	4	4	3	4	4
Q15	5	3	2	1	2	3	5	1	2	2	2	2	2
Q16	2	5	2	4	4	4	3	3	3	5	4	3	3.5

Table A1: Full Robot Acceptance Questionnaire per user. ITA is the median over the Italian users' answers; ESP is the median over the Spanish users' answers; ALL is the median over all participants.

Issue	Fix	Method
Robot moving at night	No intervention from VC between 19:00 and 9:00 Robot does not talk nor performs intervention between 22:00 and 9:00	prepilot MLS
Robot screen active at night	Robot screen turned off automatically between 22:00 and 9:00	MLS
Robot moving when user away	Testing user localization method to avoid false positives Discovery of IoT sensor fault Explanation of robot behaviour in corner cases	prepilot MLS MLS
Robot loses contact with docking while idle	Robot tries to restore contact moving forward	MLS
Robot discharging due to nav. failure	Robot tries a new docking attempt Robot turn off for safety reason when low power at night Robot tries multiple docking attempts after timeout during day Robot send "last wish" message to other component to signal issue	MLS MLS MLS MLS
HRI unable to detect speech	Robot suggests to use buttons for "Yes" and "No" instead Change of HRI dialogues to improve clarity Robot checks for connectivity and cancels intervention if issues	prepilot MLS MLS
Robot can move but not dock while its dark	Robot turns off for safety if discharging happens at night Robot turns off at critical battery level to preserve batteries Light with movement sensors placed on top of docking station to assist docking maneuver with limited visibility*	MLS MLS MLS
Unstable internet connection	Robot keepalive message to check connectivity Robot performs docking and cancels intervention if no internet connectivity Robot signals to user (voice + message on display) internet connectivity issues	MLS MLS MLS
Narrow passages (e.g., door half-closed)	Development of a navigation assistant Change topological poses location in robot map to improve navigation after inspection	prepilot MLS
Robot failing to return to docking station	Add multiple docking attempts after timeouts and failed docking Robot keepalive message to check connectivity Robot turns off for safety if discharging happens at night Robot turns off at critical battery level to preserve batteries	prepilot MLS MLS MLS
Robot lost localization	Recovery behaviour to restore localization Reinitialize localization when at docking station Robot ask to the user for help and to be moved to docking station	prepilot prepilot prepilot
Robot asks "weird" spot questions	Removed spot questions about confabulation as are signaled as "weird"	MLS
Robot performs several similar intervention in a row	Check VC for issues and bug fix	MLS
Robot stuck in a location for a few minuts	Inspect events via MLS and explain to user	MLS
Robot ask intervention too early in the morning	Add constraint to the VC policy Add redundancy costraint to the robot policy	MLS MLS
Robot WM intervention HRI inconsistent	Change of HRI dialogues of WM interaction Fix intervention scheduling time	MLS MLS
Robot CT intervention triggered every day	Check VC policy and bug fix	MLS
Robot multiple navigation failure	Inspect log and change topological poses location	MLS
Users ask to send command to robot	Users can ask to the robot to "Come here" using microphones*	MLS
Microphones trigger too many CH	Robot asks to the user to confirm CH	MLS

Table A2: A partial list of issues that emerged during the long term deployment, with the countermeasures implemented to fix them, reduce their impact, or limit their occurences. With Method, we indicate how we discovered empirically the issue in first place, either during preliminary testing (prelilot) or following the users' feedback (MLS). Items signaled as * were implemented and tested but not deployed during pilot for safety reasons during the COVID-19 pandemic, as they were designed as later improvements for the last part of the pilot.

Borghese BIO

N. Alberto Borghese graduated in electrical engineering cum laude from Milan Politecnico in 1986. He was tenured as a researcher at the Institute of Neuroscience and Bioimages of CNR, Milan in 1987 and moved to the Department of Computer Science of Università degli Studi di Milano in 2001, where he is currently Full Professor and head of the Applied Intelligent Systems Laboratory. His research activity is based on designing, developing and testing on real-problems, methods and algorithms, based on computational intelligence, with particular attention to limited processing time. In particular, he has developed novel methodology and technology in the fields of motion capture, unobtrusive tracking and sensors integration. He has also being working on developing robust, multi-scale, adaptive models for predictive regression and clustering to extract data features. More recently he has been applied this expertise to the design and realization of platforms for e-Health and e-Welfare, combining Exer-games, AI, service robots and IoT. He is co-author of more than 70 peer-reviewed journal papers, more than 100 conference reviewed papers and he holds 16 international patents.

His research has been financed significantly by industry as well as by National and European grants. In particular he has been Partner of the CNR Robocare strategic project (2001-2004), European Coordinator of the projects: FITREHAB (InterReg IVC, 2009-2011) and REWIRE (FP7, 2011-2015) and MOVECARE (H2020, 2017-2019) and partner of the ESSENCE (H2020, 2020-2022) project.

Luperto bio

Matteo Luperto is an assistant professor the Dipartimento di Informatica Giovanni Degli Antoni at the Università degli Studi di Milano (Italy). He received his Ph.D. in Information Technology from the Politecnico di Milano (Italy) in 2017. In 2012 he participated and won with the Politecnico di Milano team the Virtual Robot Competition of the RoboCup Rescue Simulation League. His research interests are semantic mapping for autonomous mobile robots in indoor environments, with a particular attention on the analysis of the structural properties of buildings, and long-term autonomy for social assistive mobile robots and their application to e-health.

Romeo BIo

Marta Romeo is a Postdoctoral Research Associate at the Cognitive Robotics Lab, Department of Computer Science, University of Manchester (UK). She is also a visiting researcher at the AI Research Center, AIST Tokyo Waterfront. She recently completed her Ph.D. under the supervision of Prof. Angelo Cangelosi on Deep Learning for Human-Robot Interaction in Elderly Care. Her research focuses on Human-Robot Interaction, with particular emphasis on evaluating the possibility of using deep learning methods to improve personalisation for Social Robots. Her current work is on Trustworthy Human-Robot Interaction and Autonomous Systems, within the UKRI TAS Node on Trust.

Borghese Photo

Luperto photo

Romeo photo

Conflict of Interest

Declaration of interests

X The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: