Toward AI-Enabled Commercial Telepresence Robots to Combine Home Care Needs and Affordability

Gloria Beraldo[1], Riccardo De Benedictis[1], Amedeo Cesta[1], Francesca Fracasso[1], Gabriella Cortellessa[1]

***Abstract*— As life expectancy increases, social and health assistance requires sustainable and affordable solutions possibly usable from one's own domestic environment. In this article, we propose a transformer-based approach combined with a *task-planning* system and enhanced with AI sub-modules to run on low-cost telepresence robots in order to support more advanced and autonomous assistance services. The proposed system allows to dynamically generate and adapt in autonomy heterogeneous robotic actions according to the information emerged during the interaction. The AI-enhanced telepresence robot was assessed in an unstructured domestic environment by 10 users. The results show an accuracy of more than 95% in the expected robot's functioning. The participants judged the system efficient, useful and intuitive, and showed a positive inclination to re-use the robot in the future. Such outcomes derive both from a proper coordination among the heterogeneous AI sub-modules in the system and from the fast capability to frequently co-adapt the interaction.**

I. INTRODUCTION

In 2022, people aged 65 years or over globally reached 771 million. The older population is projected to reach 994 million by 2030 and 1.6 billion by 2050 [1]. As life expectancy increases, there is a growing necessity and pressure to establish universal and effective health and social care systems that can foster *ageing in place*, as intended by the U.S. Centers for Disease Control and Prevention like "the ability to live in one's own home and community safely, independently, and comfortably, regardless of age, income, or ability level". Providing efficient and cost-effective home-care services will require major changes in the current ways of monitoring and delivering care services to recipients.

Research on robot-based home-care systems has shown that assistive and companion robots are promising tools to monitor, entertain, and prevent risky situations for older people [2], [3] as well as to support caregivers and healthcare professionals [4], [5]. Among them, telepresence robots appear to be more effective than other communication devices to allow people to virtually interact with others in several scenarios ranging from the remote education [6], to facilitate the museum visits [7], from older people assistance [8]–[10] to patients' support during COVID-19 pandemic [5], [11].

Although robotic platforms have advanced significantly [12], the services offered off-the-shelf for commercial telepresence robots remain quite limited to basic telepresence functions, such as bi-directional audio and video streaming and remote teleoperation capabilities, with few if any features based on the robot's autonomy. Additionally, customization possibilities are very scarce and mostly related to robot's features such as speed, color, height, camera inclination, rather than to the delivered services. This limitation makes the telepresence robots partially useful and not fully exploitable in the long-term, for instance, in the perspective of *ageing in place*. Furthermore, in this teleoperated modality, telepresence robots can surely help reduce some of the stress and mental distress experienced by caregivers [13], but a complete relief is unlikely to be achieved as the operator is engaged in the dual task of focusing on reactive handling of the robot while monitoring the older person. In addition, interacting with a person by manually maneuvering a robot could be difficult due to possible delays in the communication, especially for those who are not familiar.

The lack of effective functionalities might also affect the acceptability of keeping robots in the domestic environment by people in the long term. Indeed, people usually have high expectations about the capabilities of social robots that are expected to exhibit intelligent and adaptive behaviors according to the specific context. Therefore, robots that are not able to act as humans expect will be shelved in a corner as soon as their users are no longer motivated to use them.

The purpose of this letter is to propose a transformer-based method combined with *task-planning* and enhanced with AI sub-modules to support user's adaptive and autonomous assistance services into a low-cost commercial telepresence robot with limited computational resources. To deal with hardware limitations, the system has been properly designed focusing on advanced AI-based techniques with a threefold aim: a) enhancing the functionalities of a commercial telepresence robot to allow health monitoring and social interaction at home also when a caregiver is not present, as a complement of "traditional" assistive services; b) enabling the robot to generate heterogeneous actions in autonomy that are "context aware" without using fixed and rigid rules; c) dynamically adapting the interaction according to the situation by exploiting the information acquired and elaborated in real-time.

This letter also evaluates the functioning, robustness and user experience of the proposed AI-driven services developed inside a *three-tier architecture* into ROS and tested on a commercial robot in a domestic and unstructured environment.

II. BACKGROUND & MOTIVATION

Telepresence has been widely validated both in the lab and home settings. Several research projects investigated the potential of such devices for remotely monitoring older people and cross cultural investigations revealed a positive

---

[1] G.B., R.D.B., A.C., F.F., G.C., are with the Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy `name.surname@istc.cnr.it`

impact on users' perception as a support of independent living [8], [14]–[16], including GiraffPlus[1], SYMPARTNER[2], ExCITE[3]. These platforms have benefited both older people, and their families. In the first case, by reducing social isolation thanks to a more interactive communication with respect to phone; in the second, relatives have felt more involved and reassured by the possibility of caring their beloved ones saving travel time but still enjoying their presence while connected [4]. In some cases, such robots have reduced the effort of the formal caregivers as well [4], especially after a proper training [17]. This work investigates a possible solution for endowing an affordable commercial telepresence robot with autonomous capabilities in order to enable social and health care continuity even in case of absence of caregiver who remotely teleoperates the platform.

Previous proposals for using robotic assistive prototypes beyond the *pure telepresence* come from projects like SERROGA[4], RAMCIPR[5] and SYMPARTNER[2] [18], [19], where robotic platforms are proposed for older people who experience mild cognitive impairments and/or dementia. Although they represent a big step forward in assistive robotics in uncontrolled situations, these projects exploit custom service robots relying on advanced, robust, but expensive hardware and sensors. In fact, despite the scientific advancements, such platforms appear unaffordable to be adopted in several residential environments [20]. This work proposes to consider low cost commercial robots and to enable them to autonomously choose personalised and contextualized actions, without using logic rules to describe the interaction. The exponential number of unpredictable situations, indeed, makes it challenging to scale smoothly when formulating robot behaviors using handwritten rules. To face these challenges, a transformer-based approach is proposed which is then combined with a task-planner and enhanced with several AI sub-modules. This allows to generalize robot's actions starting from *stories* (e.g., human-understandable sequences of events and actions) thus relieving the programmer from hard-coding all the possible interactions. Thanks to the use of high-level concepts, such *stories* can also be easily defined by people without a technical background.

Another important point is associated with the personalization and the adaptation of the interaction that which is paramount for acceptance of robots [21]–[23]. Previous studies about telepresence mainly focus on the adaptation of the navigation and robot's motion (person search, detection of a fallen person, reaching a goal, etc.). This work, instead, focuses on the dynamic personalisation of interaction during cognitive and physical stimulation and dialogue consequently re-adapting the robot's activities (e.g., duration, robot's feedback, robot's dialogue, etc.) conforming to the detected state of both the person and the environment. Differently from other approaches that provide pre-model describing user preferences and habits, herein, the robot automatically infers information only by talking to people via an initial short familiarization phase.

In summary, the letter aims to investigate the following research questions (RQ):

- RQ1) Is it feasible to enhance a commercial telepresence robot with AI-based services for supporting people at home and personalise the interaction online without providing any user information in advance?
- RQ2) Are the proposed services robust and able to re-adapt according to the context when running on a commercial telepresence robot in a domestic and unstructured environment?
- RQ3) Does the interaction with the telepresence robot endowed with the proposed AI-based services have a positive impact in the subjective perception of the users?

The contribution of this letter is the concretization and implementation of the transformer-based *three-tier architecture*, originally formalized in our previous work [24], into a low-computational powered commercial telepresence robot to provide cognitive and physical exercises, reminders, and dialogue. The proposed architecture and the implemented services were tested with a feasibility study involving a group of participants out-of-the-lab.

III. An autonomous and personalised system for active and healthy ageing

A. The Commercial Telepresence Robot

The platform is Ohmni[6] (Fig. 1a), a commercial mobile robot endowed with two RGB cameras (2MP as resolution, 3.0 micron as pixel dimension, and 5865 x 3276 um as sensor's dimension) and a 2D lidar (Rplidar A1: 360 °, 12 m range, 1 °as angular resolution). This choice has a threefold motivation. First, it is suitable for domestic telepresence because of its low cost in the perspective to be affordable. Second, it does not require much space (i.e., footprint: 35.30 x 43.70 cm, height: 142 cm, folding in half) and is easily transportable. Third, it is interesting from the technical point of view, given the technological challenge of transferring advanced AI services beyond pure telepresence on a robot with limited computational resources. Ohmni has a 4 cores Intel Atom x5 z8350 processor with 1.92 GHz, 2 GB of DDR3L memory, and no GPGPU.

B. AI-Driven Services for Personalising the Interaction

This work focuses on services for fostering active ageing. Specifically, the robot acts as a coach that in autonomy can: a) suggest cognitive and physical exercises; b) guide people providing instructions and encouragements during the execution of the task; c) provide customized feedback according to their real-time performance. Furthermore, given the need to retain sharp people's memory as emerged in previous studies [18], [19], [22], the robot can set personalised reminders via voice interaction. Finally, the telepresence robot acts as a companion to engage the person in a conversation.

---

[1] `https://cordis.europa.eu/project/id/288173`
[2] `https://www.sympartner.de`
[3] `http://www.aal-europe.eu/projects/excite/`
[4] `https://www.serroga.de`
[5] `https://cordis.europa.eu/project/id/643433`
[6] `https://ohmnilabs.com`

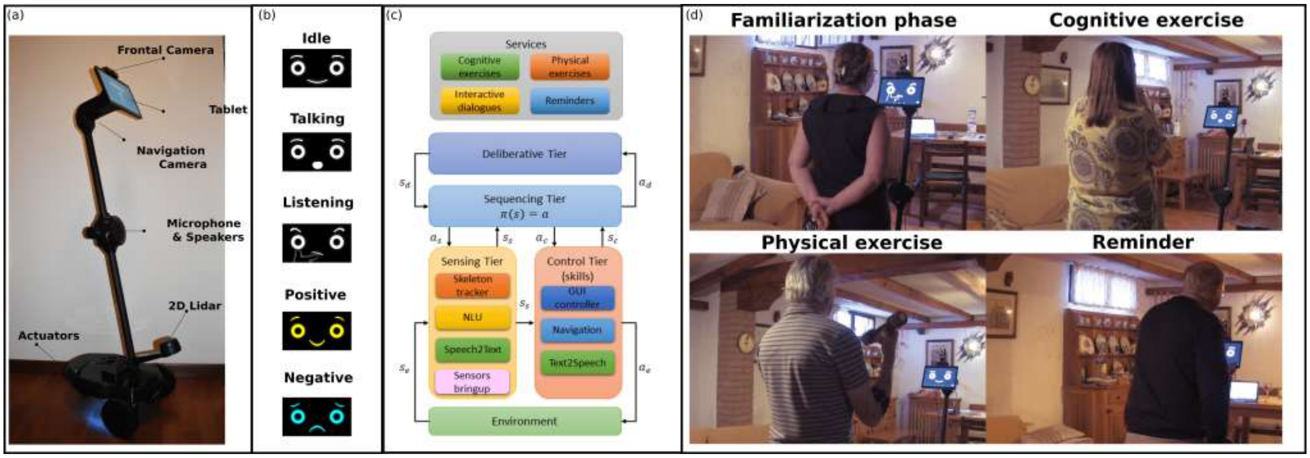


Fig. 1: (a) The commercial robotic platform used in this study. (b) The feedback displayed in the robot's screen. (c) The *three-tier* architecture behind our system. (d) The proposed AI-driven services tested in the domestic environment.

In order to realize the mentioned services the following requirements and challenges have been considered: a) to create a "ready to use" system which, thanks to an initial brief familiarization phase, dynamically collects information about the user to personalize the interaction; b) to efficiently and autonomously generate intelligent and adaptable behaviors, capable of running smoothly even on low-performance hardware; c) to simplify the characterization of the interactions, by relying on machine learning techniques, while maintaining a good control of the robot's behaviors (i.e., avoiding inconsistent actions). Indeed, it is plausible to hypothesise that the absence of personalised and automated robot's behaviors in the current commercial telepresence platforms is also associated. This work extends the concept of *stories*, namely a data representation typically used to train conversational agents to generalize unseen conversations as inside Rasa[7]. An advantage of using *stories* consists of enclosing the possible user conversations by expressing high-level concepts without specifying the specific contexts. In particular, a *story* is characterized by *intents*, *entities* and *actions*. The first ones represent the intention of the user (e.g., greetings); the second ones allows including additional information characterising the user message (e.g., location: "Rome"); the third express what the robot has to do including its answers (e.g., utter_italian_greetings). Such data are provided in input to a classifier that learns to associate unknown sentences with the same intents provided in input during the training. The work relies on the *Dual Intent and Entity Transformer (DIET)* [25], a modular transformer architecture, available inside Rasa, that allows performing both *intent classification* and *entity recognition* from expressions in natural language. This aspect is correlated to the second strength of the approach, namely the flexibility and intuitiveness in creating the training dataset. Indeed, it is possible to easily generate *stories* by a non-programmer (e.g., domain expert) since *intent, entities, and actions* are typically expressed with *high-level keywords*.

Differently from previous systems, this work exploits *stories* to manage and generalise not only conversations but all possible robot's behaviors, from robot's utterances to the feedback provided to the user via a dynamic interface in its tablet (Fig. 1b), from the generation of the assistive tasks in line with the current status of users to the robot's motions. Therefore, the examined tasks are consequently formulated - cognitive and physical exercises, dialogues, reminders - in terms of *intents*, *entities* and *actions* according to specific features of the users and the interaction like the personality, the positive and negative performance, successful/fallacious interaction (e.g., misunderstanding in the user's answers, etc.)[8]. Finally, the learned model is applied as a *policy* $\pi(s) = a$ that allows determining the next robot's actions a from the state of the system s. Additionally, the learned *policy* is combined with a *task-planner* inside a *three-tier architecture* described in the next section, that determines when co-adapting the robot's actions according to the evolution of the interaction. Indeed, the mutual interplay between an efficient *task-planner* and the learned *policy* based on *stories* allows to dynamically adapt the interaction: the first chooses the high-level robot's actions, the second characterises them based on the acquired status of the person and environment.

C. A ROS-Based Three-Tier Architecture to Generate Autonomous Robot's Behaviours

To enable autonomous robot's behaviors, the following three requirements have been considered: a) to plan a sequence of robot's tasks which, if executed, leads to goals achievement (*goal-oriented behaviors*); b) to quickly and dynamically adapt the planned tasks according to the context (e.g., due to failures, necessity to modify the duration of a task, cancel a task, add a new task); c) to perceive the context and to react to the changes which dynamically arise from the environment and the user, regardless of deliberation times (e.g., while planning or while adapting the plan). Inspired

---

[7] `https://rasa.com/`

[8] Please refer to the implementation for the complete list at `https://github.com/pstlab/sir_telepresence`

by the classical robotics [26], the proposed services were implemented inside a *three-tier architecture* that reflects such a prerequisite of *thinking slow - thinking fast* inside its structure [24], [27]. Specifically, it is composed of:

- a *deliberative tier* responsible for the generation, execution and dynamic adaptation of plans using semantic/causal reasoning;
- a *sequencing tier* executing reactive and adaptable actions determined by the learned *policy* $\pi(s)$ based on the state of the system (see Section III-B).
- a *controlling tier* in charge of converting the actions provided by the *sequencer tier* into low-level commands for the robot's actuators.

Fig. 1c shows the architecture of the system. Note that the advantages are manifold. The goals, in particular, are not entered directly into the system, but generated by the *sequencing tier*, which has the possibility of agreeing on them with the user, possibly requesting further information for their definition (and therefore carefully customizing the interaction) or, if desired, even denying them. The planned activities, aimed at achieving the generated objective, then, enter into the *sequencing tier* which, based on the dynamic evolution, can accept or, if needed, reject them, requiring further dynamic adjustments to the executing plan. Additionally, accepted activities represent one of the elements of the context that defines the interaction, allowing the *sequencing tier* to disambiguate user input based on the currently executing activity but also on previous interactions.

The *deliberative-tier* relies on a timeline-based planner, which can run on under performing hardware thanks to implemented heuristics for efficient solution generation [28] and its ability to dynamically adjust plans during execution [29]. Finally, all the three layers are implemented inside *Robot Operating System (ROS)*[9] - the standard *de facto* in robotics - to exploit the modularity and the efficient communication infrastructure provided by the middleware. Using ROS on the commercial telepresence robot is not common. In this light, the choice was also taken in order to facilitate the reproducibility of the system on other robotic platforms. To support ROS on Ohmni based on Android, specific wrappers were created for interfacing with the low-level libraries and actuators provided by the robot's producer via *Docker*[10]. To detect the execution of the *physical exercises*, the skeleton of the person detected from the robot's camera is processed and the pose variation of the joints is analysed. With regard to the *cognitive exercises*, the robot interacts with the user via dialogue. Instead, the *reminder* can be set in a hybrid way using the GUI on the robot and the voice interaction.

IV. EXPERIMENT

Based on the architecture in Section III-C, we developed different services implemented on the commercial platform Ohmni, while maintaining its original telepresence functionality. A free *dialogue service, cognitive exercises, physical exercises*, and a *reminder* service were developed. An experimental phase was conceived to assess the technical robustness of the new functionalities, as well as the user experience during the interaction.

A. Participants

This study involved 10 participants (48.40 ± 12.64 years old, four female). All participants voluntarily agreed to participate in the study and signed a written informed consent in accordance with the principles of the Declaration of Helsinki.

B. Investigated Metrics & Data Collection

Both performance and self-report metrics were considered. The investigated metrics are: a) *Accuracy*, measured as the number of expected robot's interaction over the total examined conditions; b) *Robustness*, in terms of system failures; c) *Responsiveness*, measured as the time elapsed between the current interaction and the next one; d) *User experience*, measured through an *ad hoc* questionnaire aimed at investigating aspects such as ease of use, expectations, responsiveness, usefulness, efficiency and will to use the robot in future. Table I lists the detailed questions.

TABLE I: Human-Robot Interaction a*d hoc* questionnaire. On a 5-point Likert scale where 1= Completely disagree and 5= Completely agree.

Q1. It was easy to get the information I wanted from Ohmni
Q2. It was easy to understand Ohmni in this interaction
Q3. I knew what to say or what to do at each stage of the interaction
Q4. Ohmni behaved according to my expectations during this interaction
Q5. I believe that the rhythm maintained during the interaction was appropriate
Q6. Ohmni has often been slow to respond
Q7. Ohmni was efficient in performing the task
Q8. Ohmni was useful in administering the task
Q9. I would like to use the robot in this mode in the future to perform the task

The system saves a log and a bag file per task that contain the information for the analyses. The examined data were from the robot's perception collected via onboard and external sensors (e.g., camera, laser, IMU), the utterances by the robot and the participants, the timings related to the evolution of the interaction, the planning states, the number of failures, the answers to the questionnaire. All the data were collected, processed, and stored according to the GDPR (Regulation EU 679/2016).

C. Experimental Protocol

After the introduction of the study and signing the consent form, participants were involved in: a) a *familiarization phase* composed of an emphatic dialogue based on the natural language processing, b) the execution of an illustrative *physical exercise* (i.e., biceps curls with weights), c) the execution of an exemplified *cognitive exercise* to train the sustained auditory attention (i.e., count the occurrences of a specific word in a list), d) the setting of a *reminder* of their own choice (e.g., remember to take the medicine, call the doctor, etc). The robot autonomously provided the instructions for each sub-tasks and led the interaction by

---

[9] `https://www.ros.org/`

[10] `https://www.docker.com/`

engaging the participant, contextualizing the situation based on images from the frontal camera and the sounds from microphones, providing feedback on the performances via the graphical interface on the tablet (in Fig. 1b) and through personalised dialogues. Except from phase a) which served to build up some knowledge to personalise the subsequent interactions, the order of the activities was randomised among the participants to avoid introducing confounding factors and they were established in agreement with the output from the *deliberative tier*. It is worth highlighting, that, the transition among the activities was also managed by the robot itself according to the status of the interaction. Each activity had no fixed duration since it depended on the evolution of the interaction as well as the interval time between two consecutive tasks. The user was required to confirm when starting the next one by clicking on the graphical interface on the tablet that appropriately changed according to the context as mentioned before (in this case, at the end of each task). This mechanism allows: a) to re-adapt in real-time the planned tasks by extending the duration of the current and the next activities, b) to avoid the intervention of the experimenter that could bias the entire evaluation.

The experiment was conducted in one session (i.e., one trial per task) in an ecological domestic environment represented in Fig. 1d, naturally affected by noise and low-speed connection, where the enhanced telepresence robot was free to operate without the intervention of the operator. The user could freely choose where to place in front of the robot. However, while performing the physical exercise, participants were required to keep a distance of at least 1m from the robot to guarantee the presence of the examined human body parts in the field of view of the camera.

D. Hardware & Tools for the Experiment

Besides the robot, to assess the responsiveness in detecting the repetitions based on the *skeleton tracker* during the *physical exercise*, an Inertial Measurement Unit (IMU) was used to compute offline the time needed by the implemented service, that processes the keypoints of the shoulder, elbow, and wrist for each arm and classifies the motion in two main phases: "outstretched arms" and "bent arms" based on the angles among these joints and the synchrony of the arms, verifying the variation in the orientation measured by the sensor. To this purpose, humans wore a non-invasive wearable bracelet (Myo Armband for its ease of use and integration) during the physical exercise service. Furthermore, the *skeleton tracker* applies *Mediapipe*[11] given the possibility to run it on devices with limited computation resources as the selected telepresence robot. The *text-to-speech* and *speech-to-text* exploit the *Google services* on the cloud.

V. RESULTS

All the participants completed all the tasks, but one of them partially finished the reminder task because of robot's connectivity technical problems (i.e., independent from the developed services). The distribution of robot's and participants' utterances can be seen in Fig. 2a. Focusing only on the human-robot interaction, the experience lasted on average 879.09s (i.e., 14 minutes and 39 seconds), which are subdivided into the four examined tasks in Fig. 2b.

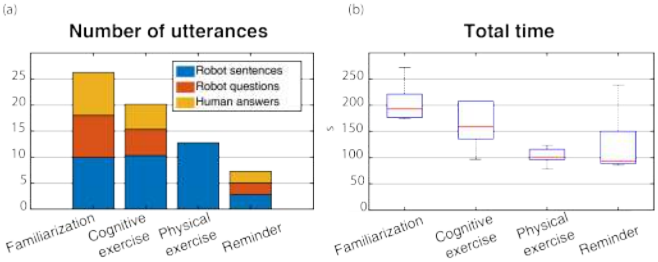


Fig. 2: (a) Number of utterances during human-robot interaction. (b) The time distribution of each task.

A. Accuracy, Robustness & Responsiveness

Overall, the proposed AI-based services performed correctly as emerged from the very low percentage of failures in Fig. 3 – below the 5% – during the whole interaction. This result suggests that on the one side, the designed services work in line with the expected robot's behaviors (e.g., few failures by the robot), on the other, they are intuitive for the participants (e.g., few failures by humans).

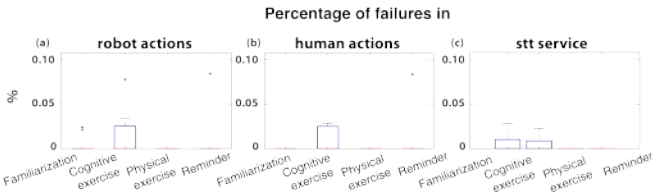


Fig. 3: (a) Failures in the robot actions. (b) Failures in the human actions. (c) Failures in the *speech-to-text (stt)* service.

In order to show the proper coordination and the execution of heterogeneous sub-modules of the system, results are presented on the physical exercise task using the IMU signals along the three axes as reference. Specifically, the *skeleton tracker* output is examined based on the robot's camera images; the detected repetitions with respect to the peaks in the IMU signals are evaluated in the corresponding temporal windows and the related vocal feedback provided by the robot are examined. Fig. 4a shows an example of the resulted association among the IMU signals peaks per axis (in blue), the related detected repetitions based on the skeleton (green dashed lines) and the robot's vocal feedback (orange dot-dashed lines). To evaluate the responsiveness of the system, the following variables are computed: delays between the execution of the movements from the IMU data, the detection of the repetitions from the *skeleton processing* and the related vocal notifications. Results are represented in Fig. 4c-d. It is worth noting that the time difference between the execution of the human repetition and its detection from the skeleton is very short, namely on average 0.46 ±0.62s as shown in

---

[11] `https://developers.google.com/mediapipe`

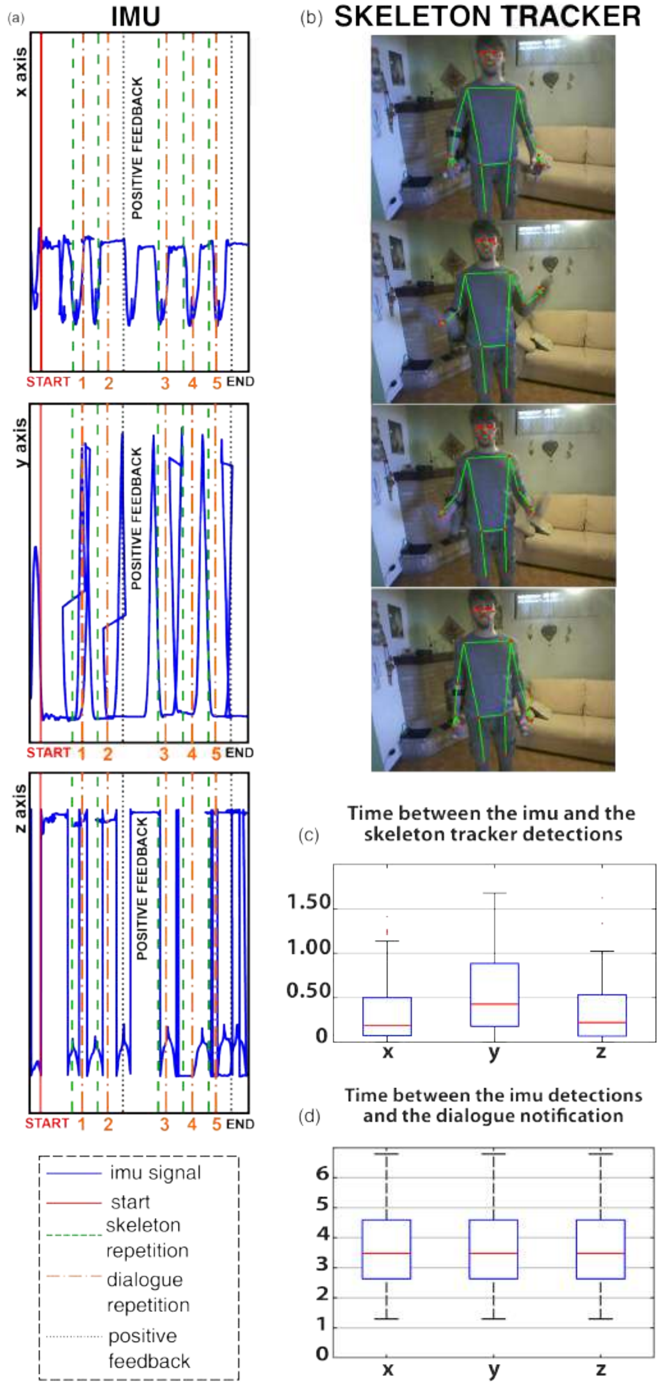


Fig. 4: (a) Representation of the IMU signal and the repetitions detected from the *skeleton tracker* (i.e., green dashed lines) and the notification from the dialogue (i.e., orange dot-dashed lines). (b) Example of a repetition from the *skeleton tracker* side. (c) The time between the IMU and the *skeleton tracker* detection of repetitions over subjects. (d) The time between the IMU and the dialogue notification over subjects.

Fig. 4c. Instead, as it appeared in Fig. 4d, the *text-to-speech* takes about 8 times longer than the *skeleton processing* before providing the vocal feedback. Since the vocal processing is made on cloud, as mentioned in Section III-C, such a result is affected by the network speed that was on average 4.56 Mbps in download and 0.16 Mbps in upload[12].

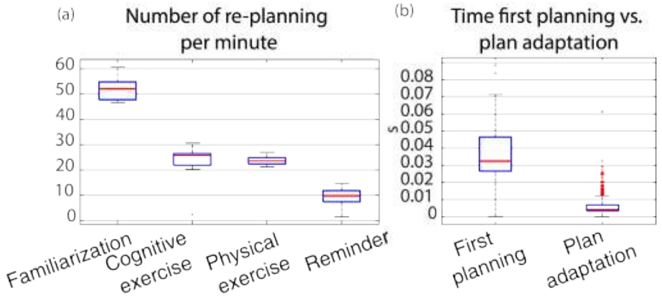


Fig. 5: (a) The number of co-adaptations of the original plan per each examined task. (b) Comparison of the times required for planning the first time (e.g., at the beginning) vs. for co-adapting the plan during the interaction.

B. Co-Adaptation of the Interaction over the Task

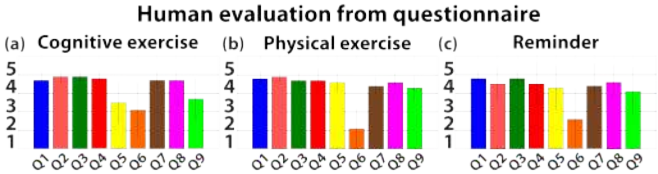
The evolution of the personalised interaction in terms of plan co-adaptation over the task was also assessed. As expected, the original plan and the schedule found at the beginning are highly modified according to the participants' answers and their performance in the exercises (e.g., continuous co-adaptation) in Fig. 5a, even after the *familiarization* phase. Considering this aspect and the necessity to guarantee the coordination among the heterogeneous modules to avoid introducing additional delays due to the re-synchronization of the planner, it is worth highlighting that, the time required by the system to co-adapt the plan represented in Fig. 5b is extremely short (i.e., on average $5.20 \pm 3.40 \times 10^{-3}$ s) and significantly less compared to the creation of a new plan from scratch considering the changes in the context. The time-saving percentage in co-adapting the plan was 87.20%. This result confirms the advantage of having an autonomous system that contextualizes the situation and personalises the interaction proactively by re-exploiting the computation made at the beginning with the introduction of new constraints (e.g., temporal, add/remove sub-goals), avoiding solving a new problem from scratch and/or involving an external operator to manually drive the robot.

C. User Experience Evaluation

Results on user experience are displayed in Fig. 6 showing the average and the standard deviation for each item of the questionnaire. In all three tasks, participants retrieved the necessary information easily, understood the robot easily, and knew what to say during the interaction, suggesting that the proposed system is intuitive in each of the three experimented services. Furthermore, in line with such results, the robot's behaviors in all the tasks matched the human's expectations. Regarding responsiveness, some differences between the three services can be noticed. Specifically, participants consider the rhythm of the interaction to be more appropriate during the *physical exercise* and the *reminder* setting than

---

[12] `https://speed.measurementlab.net/`



task	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
cognitive exercise	4.70 ±0.46	4.90 ±0.30	4.90 ±0.30	4.80 ±0.40	3.50 ±1.11	3.10 ±1.45	4.70 ±0.46	4.70 ±0.46	3.70 ±1.10
physical exercise	4.80 ±0.40	4.90 ±0.30	4.70 ±0.46	4.70 ±0.64	4.60 ±0.49	2.10 ±1.04	4.40 ±0.49	4.60 ±0.49	4.30 ±0.64
reminder	4.80 ±0.40	4.50 ±0.92	4.80 ±0.4	4.50 ±1.20	4.30 ±1.19	2.60 ±1.62	4.40 ±1.20	4.60 ±0.91	4.10 ±1.30

Fig. 6: Results from the questionnaire related to (a) cognitive exercise, (b) physical exercise, (c) reminder tasks. Mean and standard deviation (Mean ± SD) are reported.

during the *cognitive exercise*. Coherently, the robot was perceived as a bit slow while providing the cognitive exercise than the two other tasks. However, overall, the robot was considered efficient and useful (see scores from Q7 and Q8). Such results were verbally confirmed by some participants at the end. Finally, participants appear to be well inclined to re-use the robot in the future, especially for helping them during *physical exercises* and *reminders*.

VI. DISCUSSION

This work proposes a transform-based method to enhance a commercial telepresence robot with a *task-planner* and AI-based services for delivering personalised cognitive and physical exercises, reminders, and interactive dialogue without the need to necessarily exploit the teleoperation. Interaction is managed by the robot that autonomously proposes the examined tasks, acquires information to provide personalised feedback without relying on a profile given in advance and re-adapts its behaviors in real-time. The robot was able not only to co-adapt the interaction dynamically even after a short dialogue, but also to guarantee a good level of robustness. Results reported in Section V-A demonstrated that the total failures were very few. System performance results are also confirmed by the self-report questionnaire: participants considered definitely easy to understand the robot and get information from it; they always knew what to say during the interaction and judged the robot's behaviour plausible and efficient. Considering research question RQ1, the accuracy suggests that it is possible to successfully augment the capabilities of a commercial telepresence robot with AI-based services. Beyond the correct functioning of the system, responsiveness was also assessed that plays an important role towards the acceptability. This aspect is not trivial since the system required the coordination of different modules running on a commercial telepresence robot with limited resources and tested in a domestic scenario with a low-speed connection. To facilitate the fulfillment of the requirement, the system was deployed into a *three-tier architecture* that includes the capability of *thinking-slow* associated with cognitive reasoning based on heuristics and *thinking-fast* to ensure quick adaptation. In addition, standard and optimized communication infrastructure provided by ROS was exploited, also with the intent to facilitate the deployment of the system on other robotic platforms. To analyse the coordination, *physical exercise* were considered that requires the activation and synchronization of the largest number of sub-modules and for which reference of the real timings from the IMU signals acquired with an external sensor were available. Fig. 4 shows the expected progress of the interaction highlighting fast processing of the skeleton, while the *speech-to-text* and *text-to-speech* required more time. These observations are in line with expectations since the first one runs onboard the robot, while the second need access to the cloud. As emerged in other tests [17], this delay is due to the network speed, which still represents one of the main barriers to the regular use of such platforms [20]. To evaluate the suitability of the rhythms of the interaction, opinions of participants were elicited. They judged the time of the interaction as appropriate, especially for the physical exercise. Nonetheless, it seems that there is a difference among the tasks, in particular, the cognitive exercise resulted more affected by the timings required by the *speech-to-text* and *text-to-speech*, probably because it is completely administrated via dialogue.

The second key objective of this work was to enable the robot to personalise and re-adapt the interaction according to the specific situation. Anticipating all possible interactions without the generalization capabilities of machine learning, indeed, makes logic rule-based methods difficult to scale as the number of interactions increases. To face this challenge, *in primis*, an extension of use of *stories* was proposed for handling all the robot's actions, taking inspiration from the approach behind Rasa, given the simplicity in generating data to learn in natural language and the potentiality of classifying unseen data to the same high-level intents by leaving out the specific knowledge. A *deliberative-tier* was the included, relying on timeline-based planning, that tries to adjust the original plan to the situation by re-using the information derived from the first planning without the necessity to re-plan every time from scratch. Indeed, as emerged in Fig. 5, there are several factors influencing the interactions that require to modify the initial plan very frequently. As for research question RQ2, the system revealed to tolerate a reiterated number of re-planning that allows a continuous and successful adaptation made in autonomy by the system thanks to the fast re-planning based on the computations from the previous planning. Finally, RQ3 is focused on the user experience. Overall, the results from questionnaires in Fig. 6 show a positive experience. Participants recognized the utility of the system and manifested the will to re-use the robot.

VII. CONCLUSION & FUTURE WORKS

This work faces the technological challenge of supporting advanced and affordable active-ageing services as a complement to those offered via the traditional social-health system on a commercial telepresence robot with limited computational resources in a domestic environment. From the scientific point view, it showed that the mutual cooperation between a *task-planner* based on heuristics and a new transformer-based method enhanced with heterogeneous AI-driven services allows handling and personalizing the interaction in autonomy, without relying on *a priori* user's model nor on *ad hoc* logic rules beyond the hardware limitations. The AI-enhanced telepresence robot was able to properly co-adapt the interaction through a reliable communication among the modules in the ROS *three-tier architecture* and received a positive consensus from participants. One limitation of the current work is the reduced sample size of participants which prevents from obtaining definitive results. However, this study was designed upstream as a preliminary assessment of the system functioning and timings in an more unstructured environment with respect to the laboratory settings, before proceeding with a more intensive evaluation of the technology efficacy and acceptance. Future directions will also investigate new learning techniques for customizing the interaction that consider the hardware constraints. Thanks to the high modularity of the system, the AI-based services will be distributed and tested also according to the recent *fog-edge-cloud* paradigm to process the sensible information inside local *fog* nodes reducing the timings for data processing and increasing the privacy protection.

ACKNOWLEDGMENT

Authors thank the participants and colleagues from the University of Genova to design the robot's expressions. The research was funded by "SI-Robotics: Social robotics for active and healthy ageing" (Italian M.I.U.R., PON—Ricerca e Innovazione 2014-2020—G.A. ARS01 01120), Cleverness (FOE 2020) (DUS.AD016.133) and FOCAAL projects.

REFERENCES

[1] V. Gaigbe-Togbe, L. Bassarsky, D. Gu, and a. L. Z. Thomas Spoorenberg, "World population prospects 2022: Summary of results," United Nations Department of Economic and Social Affairs, Population Division, Tech. Rep., 2022.

[2] G. Mois and J. M. Beer, "The role of healthcare robotics in providing support to older adults: a socio-ecological perspective," Current Geriatrics Reports, 2020.

[3] M. E. Pollack, "Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment," AI Magazine, vol. 26, no. 2, p. 9, 2005.

[4] M. Niemelä, L. Van Aerschot, A. Tammela, I. Aaltonen, and H. Lammi, "Towards ethical guidelines of using telepresence robots in residential care," International Journal of Social Robotics, vol. 13, no. 3, pp. 431–439, 2021.

[5] B. Isabet, M. Pino, M. Lewis, S. Benveniste, and A.-S. Rigaud, "Social telepresence robots: A narrative review of experiments involving older adults before and during the covid-19 pandemic," International Journal of Environmental Research and Public Health, vol. 18, no. 7, p. 3597, 2021.

[6] P. Thompson and S. Chaivisit, "Telepresence robots in the classroom," Journal of Educational Technology Systems, p. 00472395211034778, 2021.

[7] E. Chang, "Museums for everyone: Experiments and probabilities in telepresence robots," in Exploring Digital Technologies for Art-Based Special Education. Routledge, 2019, pp. 65–76.

[8] A. Cesta, G. Cortellessa, F. Fracasso, A. Orlandini, and M. Turno, "User needs and preferences on AAL systems that support older adults and their carers," J. Ambient Intell. Smart Environ., vol. 10, no. 1, pp. 49–70, 2018.

[9] S. Laniel, D. Létourneau, F. Grondin, M. Labbé, F. Ferland, and F. Michaud, "Toward enhancing the autonomy of a telepresence mobile robot for remote home care assistance," Paladyn, Journal of Behavioral Robotics, vol. 12, no. 1, pp. 214–237, 2021.

[10] K. Winterstein, L. Keller, K. Huffstadt, and N. H. Müller, "Acceptance of social and telepresence robot assistance in german households," in International Conference on Human-Computer Interaction. Springer, 2021, pp. 326–339.

[11] C. Esterwood and L. Robert, "Robots and COVID-19: Re-imagining human–robot collaborative work in terms of reducing risks to essential workers," SSRN 3767609, 2021.

[12] B. Isabet, M. Pino, M. Lewis, S. Benveniste, and A.-S. Rigaud, "Social Telepresence Robots: A Narrative Review of Experiments Involving Older Adults before and during the COVID-19 Pandemic," International Journal of Environmental Research and Public Health, vol. 18, no. 7, 2021.

[13] M. Pérez-Cruz, L. Parra-Anguita, C. López-Martínez, S. Moreno-Cámara, and R. del Pino-Casado, "Burden and anxiety in family caregivers in the hospital that debut in caregiving," International Journal of Environmental Research and Public Health, vol. 16, no. 20, p. 3977, 2019.

[14] J. M. Beer and L. Takayama, "Mobile remote presence systems for older adults: acceptance, benefits, and concerns," in Proceedings of the 6th international conference on Human-robot interaction, 2011, pp. 19–26.

[15] A. M. Seelye, K. V. Wild, N. Larimer, S. Maxwell, P. Kearns, and J. A. Kaye, "Reactions to a remote-controlled video-communication robot in seniors' homes: a pilot study of feasibility and acceptance," Telemedicine and e-Health, vol. 18, no. 10, pp. 755–759, 2012.

[16] J. Gonzalez-Jimenez, C. Galindo, and C. Gutierrez-Castaneda, "Evaluation of a telepresence robot for the elderly: a spanish experience," in International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, 2013, pp. 141–150.

[17] L. Fiorini, A. Sorrentino, M. Pistolesi, C. Becchimanzi, F. Tosi, and F. Cavallo, "Living with a telepresence robot: results from a field-trial," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 5405–5412, 2022.

[18] H.-M. Gross, S. Mueller, C. Schroeter, M. Volkhardt, A. Scheidig, K. Debes, K. Richter, and N. Doering, "Robot companion for domestic health assistance: Implementation, test and case study under everyday conditions in private apartments," in 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2015, pp. 5992–5999.

[19] H.-M. Gross, A. Scheidig, S. Müller, B. Schütz, C. Fricke, and S. Meyer, "Living with a mobile companion robot in your own apartment-final implementation and results of a 20-weeks field study with 20 seniors," in 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 2253–2259.

[20] M. Wang, C. Pan, and P. K. Ray, "Technology entrepreneurship in developing countries: Role of telepresence robots in healthcare," IEEE Engineering Management Review, vol. 49, no. 1, pp. 20–26, 2021.

[21] C. Moro, G. Nejat, and A. Mihailidis, "Learning and personalizing socially assistive robot behaviors to aid with activities of daily living," ACM Transactions on Human-Robot Interaction (THRI), vol. 7, no. 2, pp. 1–25, 2018.

[22] C. Di Napoli, G. Ercolano, and S. Rossi, "Personalized home-care support for the elderly: a field experience with a social robot at home," User Modeling and User-Adapted Interaction, pp. 1–36, 2022.

[23] A. Umbrico, A. Cesta, G. Cortellessa, and A. Orlandini, "A holistic approach to behavior adaptation for socially assistive robots," International Journal of Social Robotics, vol. 12, no. 3, pp. 617–637, 2020.

[24] R. De Benedictis, G. Beraldo, G. Cortellessa, F. Fracasso, and A. Cesta, "A transformer-based approach for choosing actions in social robotics," in International Conference on Social Robotics. Springer, 2022.

[25] T. Bunk, D. Varshneya, V. Vlasov, and A. Nichol, "DIET: lightweight language understanding for dialogue systems," CoRR, vol. abs/2004.09936, 2020.

[26] E. Gat, "On Three-Layer Architectures," in Artificial Intelligence and Mobile Robots. AAAI Press, 1997, pp. 195–210.

[27] D. Kahneman, Thinking, fast and slow. Macmillan, 2011.

[28] R. De Benedictis and A. Cesta, "Lifted heuristics for timeline-based planning," in ECAI 2020. IOS Press, 2020, pp. 2330–2337.

[29] R. De Benedictis, G. Beraldo, A. Cesta, and G. Cortellessa, "Incremental timeline-based planning for efficient plan execution and adaptation," in AIxIA 2022 – Advances in Artificial Intelligence. Springer, 2022.