GRANT AGREEMENT N. 871245

Deliverable D1.5

User feedback from the intermediate validation (realistic/relevant environments)

Due Date: 24/05/2023
Main Author: AP-HP
Contributors: ERM, INRIA
Dissemination: Public Deliverable

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

DOCUMENT FACTSHEET

Deliverable no.	D1.5: User feedback from the intermediate validation (realistic/relevant environments).
Responsible Partner	ERM
Work Package	WP1: Experimental Validation
Task	T1.4: Intermediate Experimental Validation
Version & Date	24/05/2023
Dissemination level	[X] PU (public) [ ] CO (confidential)

CONTRIBUTORS AND HISTORY

Version	Editor	Date	Change Log
1	AP-HP	24/05/2023	First Version
Final	ERM, AP-HP	31/05/2023	Final version

APPROVALS

Authors/editors	Maribel Pino (AP-HP) Sébastien Dacunha (AP-HP) Anne-Sophie Rigaud (AP-HP) Lauriane Blavette (AP-HP) ERM
Task Leader	ERM
WP Leader	ERM

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Contents

1. General framework for the SPRING project assessments 4
2. Description of the evaluation environments and research procedures in each study setting 5
2.1 Study 1: Simulated relevant environment (Hospital Living Lab environment) 5
2.2 Study 2: Real environment study (Geriatric Day Care Hospital - realistic environment) 9
2.3 Software and hardware architecture evolution 12
3. Description of the methodology and the results of three waves of intermediate evaluations 14
3.1 Summary of the assessments 14
3.2 Specific description of each of the studies conducted in a simulated relevant environment (Study 1) 15
3.2.1 Study 1: Wave 1 experiment (Living Lab - Simulated environment) 15
3.2.2 Study 1: Wave 2 experiment (Living Lab - Simulated environment) 31
3.2.3 Study 1: Wave 3 experiment (Living Lab - Simulated environment) 55
4. General discussion and conclusions of the intermediary assessments 80
5. References 84
6. Annexes 85
Annex 1: System Usability Scale SUS, adapted French version 85
Annex 2. Acceptability E-Scale AES, adapted version 87
Annex 3. Observation grid used during the interaction 89
Annex 4. Guide for the semi-structured interviews (ethical dimension) 91
Annex 5. Example Task Sheet 92
Annex 6. Appreciation of the dialogue capabilities of the robot (HWU questionnaire) 93

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Executive summary

Deliverable D1.5 presents the results of a series of intermediary experiments conducted by AP-HP in relevant environments. Different versions of the SPRING system were assessed with its respective modules integrated (ERM).

The experiments were held in the Living Lab of Broca Hospital (AP-HP) where volunteers were recruited to participate in the experiments.

The document contains four sections:

1. Description of the general framework for the SPRING project experiments in both relevant and realistic environments
2. Description of the evaluation environments and research procedures (Hospital Living Lab and Day Care Hospital consultations)
3. Description of the methodology and the results of three waves of intermediate evaluations
4. General discussion and conclusion of the intermediary assessments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

1. General framework for the SPRING project assessments

Within the framework of the SPRING project, five use cases for a socially assistive robot in a hospital context are developed.

a) The robot provides a reception or welcoming of the hospital users
b) The robot delivers information and reminders on how to have safe social interactions
c) The robot provides assistance to patients throughout the care process
d) The robot provides orientation and guidance related to the hospital services
e) The robot provides entertainment activities for persons who attend the hospital

With the aim of assessing the relevance and effectiveness of these use cases, we are conducting experimental research involving human subjects in relevant and realistic environments. Two experimental frameworks were therefore defined to meet the needs of the project (Figure 1). These frameworks are described in the following section (Section 2).

Figure 1. Global description of the experimental assessment for the SPRING project

This report describes the main results from the **intermediate experiments** that have been conducted between April 2022 and March 2023 within the framework of the SPRING project in a simulated-relevant environment (Study 1). User feedback from the realistic environment (Study 2) will be reported in deliverable *D1.6: User feedback from the final validation (relevant environments)* using results from the ongoing final wave of experiments.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

2. Description of the evaluation environments and research procedures in each study setting

2.1 Study 1: Simulated relevant environment (Hospital Living Lab environment)

Objectives

The Living Lab framework (simulated environment) allows testing with older adults and care professionals the robot's functionalities and the use cases respond to three specific requirements of the project:

(a) To collect user and environmental data (voice, video, maps) that facilitates the handling of the robot and the learning that the robot achieves relative to the environment (Figure 2).

(b) To evaluate the progress of the SPRING system and the new functionalities that are gradually being developed by the project partners,

(c) To gather information for future improvements of the different modules of the system and its integration, as well as to identify upgrade needs.

The confrontation of the robot with potential users in these experiments also allows a better understanding of the determinants of acceptance of socially assistive robots by users.

Figure 2. Robot ARI in the Living Lab setting

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Description of the environment

The **Broca Living Lab** is a research and innovation organization based in **Broca Hospital (AP-HP, Paris)**, and a pioneer of the living lab approach in France. Its main objective is to promote the development and use of effective and ethical technology-based solutions aiming at improving the autonomy and quality of life of older adults and supporting their informal and formal caregivers in their tasks. Within this framework, research work is developed in close collaboration with patients, their families, academic partners, industrial partners, and public and private funders.

In order to develop technology-based that truly meets the needs of users, the Living Lab approach uses co-design methodologies, involving all stakeholders at all stages of the design process: from the initial needs analysis to the testing of successive prototypes, until a solution is ready to be used and well accepted by its intended public.

The Living Lab has an easily configurable architecture that can be adapted to the needs of each project. The facility allows for the observation of user behavior in a simulated environment, under controlled conditions, and in complete safety. This setting supports the study of target users' interactions with technological devices using non-intrusive methods (audio and video recordings, behavioral analysis, etc.) (Figure 3).

Figure 3. Configuration of the Living Lab environment

Modalities of the experiments

The studies in the simulated environment are conducted using a system of evaluation campaigns, or **waves.** These testing waves are organized with all the project partners, generally following immediately a phase of integration of the robot's technical modules. At the moment of drafting this document, three evaluation waves have taken place in the living lab:

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

- **Wave 1:** April 2022
- **Wave 2:** October 2022
- **Wave 3:** March 2023

Participants

For Study 1, eligible individuals are those who wish to participate in the study. They must have expressed no objection to participation in the study and meet the inclusion criteria defined below:

- ***Older adults***: be over 60 years old; be sufficiently fluent in French to understand simple instructions, most volunteers are hospital outpatients or informal caregivers of patients followed in the hospital, but who participate in the SPRING evaluation outside of the hospital consultation.

General procedure of the tests

After having been informed about the assessment procedure and having signed the written consent to participate in the assessment session:

- Participants are invited to the Living Lab of the Broca Hospital for a testing session (length between 1-2 hours)
- Individual and collective sessions are possible to schedule according to the assessment needs (collecting data regarding navigation, interaction, individual conversations, multi-party conversations, engagement analysis, etc.)
- Participants are invited to interact with the robot and test a use case defined in advance (the interaction is video recorded)
- At the end of the interaction, a mixed-methods assessment takes place. The participants fill out some questionnaires (quantitative approach with standardized scales) and take part in an interview (qualitative approach).

Main dimensions of the assessment

- Quality of **human-robot interaction :** user satisfaction that the robot meets the needs in the five use cases

- **Usability**: degree of efficiency in using the robot to achieve a specific goal in a particular context

- **Acceptability**, understanding the willingness of users to adopt the robot in their social and professional environment

- **Ethical** issues generated by the presence of the robot (understand the determinants of rejection or adoption).

Assessment instruments

(a) Usability: " System Usability Scale" (Bangor et al., 2008; Brooke, 1996). This is a 10-item questionnaire reliable for measuring usability. *Usability* has been defined as the ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment; where 'ease of use' affects user performance and satisfaction, and 'acceptability' affects whether or not the product is used (Bevana et al., 1991).

The advantages of the SUS instrument are that it is a very easy scale to administer to participants, it can be used on small samples with reliable results, and it can effectively differentiate between usable and unusable systems or products. The SUS score is located on a scale ranging from 0 to 100, a SUS score above 68 would be considered above average and anything below 68 is below average. Adjective ratings provide an interpretation of the SUS score (Figure 4). The scale is available in Annex 1.

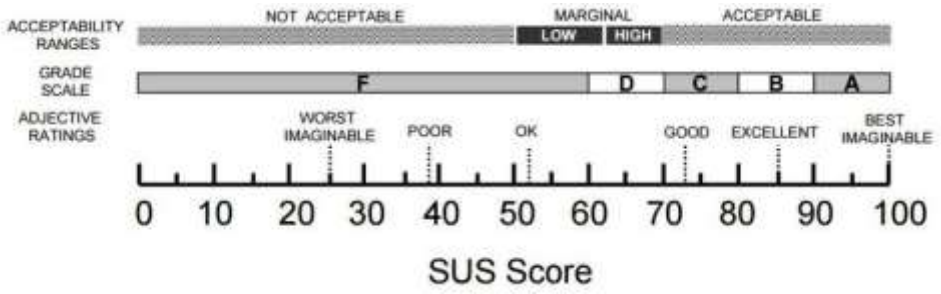


Figure 4. Overall System Usability Scale score

(b) Acceptability: "Acceptability E-Scale" (AES), French version (Micoulaud-Franchi et al., 2016), adapted for use with a robot. Acceptability is defined as the psychological antecedents of the behavioral intention to use technology without experience of the system. The AES is a 6-item self-reported questionnaire that evaluates the extent to which users find a system acceptable focusing on two dimensions: usability and satisfaction (Hayotte et al., 2020). Scale total scores range from 6 to 30 (See Annex 2)

(c) Interaction data (comments and observations): used to provide info for partners (voice recognition, conversation, behavior, movements, …) using video recordings + observation grid (See Annex 3).

(d) Ethical aspects: Semi-structured interviews built from existing ethical questionnaires (Alaiad & Zhou, 2014) and the "Ethical Acceptability Scale" (Peca, 2016) used to explore general ethical

issues related to the use of social and assistive robots in healthcare environments, their advantages, disadvantages, and risks as perceived by potential users (See Annexe 4).

Data analysis conducted

To have a holistic understanding of the research topic, the data collected is qualitative and quantitative. For this intermediate report, quantitative data (gathered with the assessment scales) is analyzed using descriptive and inferential statistics (i.e., means, standard deviations, frequencies, etc.). Qualitative data (gathered during the semi-structured interviews) is analyzed using inductive thematic analysis.

2.2 Study 2: Real environment study (Geriatric Day Care Hospital - realistic environment)

Objectives

Study 2 aims primarily at assessing the performance, acceptability, and usability of the SPRING robot among a sample of users (patients and accompanying persons) in a real-life environment (Geriatric Day Care Hospital) (expected outcome n=100 patients, 100 accompanying persons). The second objective is to identify ethical issues linked to the use of socially assistive robots in healthcare settings.

Based on the results of this study, we will determine the possibilities for socially assistive robotics to improve the quality of patient care in the hospital and the work of healthcare professionals.

Description of the environment

The Broca hospital (AP-HP) is a geriatric day care hospital, aimed at providing specialized care and support for older adults, who may have physical or cognitive impairments. The Broca day hospital, therefore, provides a range of geriatric consultations (neurology, oncology, cardiology, behavioral disorders/psychiatry, memory consultations, etc.)

Participants

For Study 2, eligible individuals (patients of the geriatric service and accompanying persons) are those who wish to participate in the study. They must have expressed no objection to participation in the study and meet the inclusion criteria defined below.

Group 1: Older adults - patients

- Patients attending one of the Broca Day Care Hospital consultations

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

- Being over 65 years old
- Have a sufficient understanding of French
- Without distinction of gender, socio-professional category or ethnic origin
- Without severe cognitive disorders (MMSE[1] > 10) or symptoms of reality alteration (delirium, hallucination)
- Having expressed no opposition to participate in the study

Group 2: Accompanying persons

- Accompanying family members or friends (informal caregivers). Primarily spouses (older adults) or children of patients, also family friends or neighbors
- Being over 18 years old
- Have expressed non-opposition to participating in the study

General procedure of the tests

After having been informed about the assessment procedure and having signed the written consent to participate in the assessment session:

- Participants are invited to a specific room of the Day Care Hospital for 30-40 min.
- Individual & patient-companion sessions only
- Participants are invited to interact with the robot and test a use case defined in advance (video-recording)
- At the end of the interaction, each participant separately fills out questionnaires (mixed, quantitative approach with standardized scales) and separately takes part in an interview.

Assessment instruments

(a) Usability: " System Usability Scale" (Bangor et al., 2008; Brooke, 1996). This is a 10-item questionnaire reliable - for measuring usability. *Usability* has been defined as the ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment; where 'ease of use' affects user performance and satisfaction, and 'acceptability' affects whether or not the product is used (Bevana et al., 1991).

The advantages of the SUS instrument are that it is a very easy scale to administer to participants, it can be used on small sample sizes with reliable results, and it can effectively differentiate between usable and unusable systems. The SUS score is located on a scale ranging from 0 to 100, a SUS score above 68 would be considered above average and anything below 68 is below average. Adjective ratings provide an interpretation of the SUS score (Figure 4). The scale is available in

[1] The Mental-State Examination (MMSE) is a commonly used set of questions for screening cognitive function that can be used to indicate the presence of cognitive impairment. Scores of 25-30 out of 30 are considered normal, 21-24 as mild, 10-20 as moderate and below 10 as severe impairment.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Annex 1.

(b) Acceptability: "Acceptability E-Scale" (AES), French version (Micoulaud-Franchi et al., 2016), adapted for use with a robot (Annex 2). Acceptability is defined as the psychological antecedents of the behavioral intention to use technology without experience of the system. The AES is a 6-item self-reported questionnaire that evaluates the extent to which users find a system acceptable focusing on two dimensions: usability and satisfaction (Hayotte et al., 2020).

(c) Interaction data (comments and observations): used to provide info for partners (voice recognition, conversation, behavior, movements, …) using video recordings + observation grid (See Annex 3).

(d) Ethical aspects: Semi-structured interviews built from existing ethical questionnaires (Alaiad & Zhou, 2014) and "Ethical Acceptability Scale" (Peca, 2016) used to explore general ethical issues related to the use of social and assistive robots in healthcare environments, their advantages, disadvantages, and risks as perceived by potential users (See Annexe 4).

Data analysis conducted

To have a holistic understanding of the research topic, the data collected is qualitative and quantitative. For this intermediate report, quantitative data (gathered with the assessment scales) is analyzed using descriptive and inferential statistics (i.e., means, standard deviations, medians percentages, Kruskal-Wallis, and Mann-Whitney tests). Other statistical methods will be used for the final analysis, once the assessments will be finished towards the end of the project. Qualitative data (gathered during the semi-structured interviews) is analyzed using thematic analysis.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

2.3 Software and hardware architecture evolution

Wave 1: Wizard Of Oz mode

As the robot did not yet have any modules from the project partners, the robot was controlled by an operator (Wizard of Oz) through a tablet.

In this wave, there was no program in the robot. The tablet just served as a graphical interface.

The Wizard of Oz program is run on a computer. It receives commands from the tablet and then transfers the commands to the robot. It also reports the status of the robot so that the operator can see if the robot is operational.

The tablet interface is a series of buttons, each with a sentence for the robot to say.

In addition to speech control, there are also buttons to record the robot's sensors. The recorded data is stored on a NAS server.

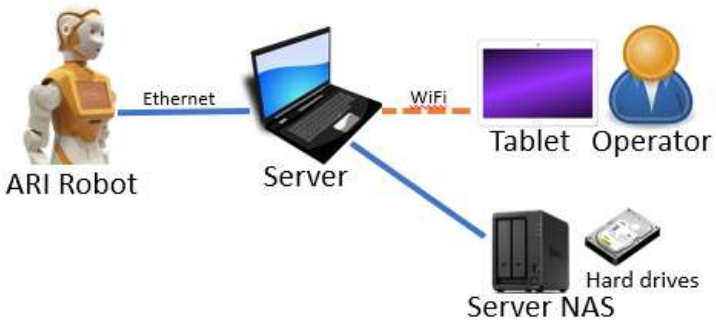


Diagram of the communications used for wave 1

Wave 2: Integration of the partners' modules

The project partners installed their programmes on a computing server, with the exception of the cartography, which was installed in the robot, and the chatbot, which remained on the HWU Cloud.

Now the robot is able to detect people, locate them, move around, hold a conversation and display the transcript on its on-board screen.

The robot itself still does not make any decisions, it is the partners' modules installed on the server that drive the robot. There is no longer an operator.

The network architecture has evolved. A powerful computing server is needed to run the partner modules. A supervisory computer was used to monitor the correct operation of the system. A secure internet connection was also added to connect to the HWU chatbot.

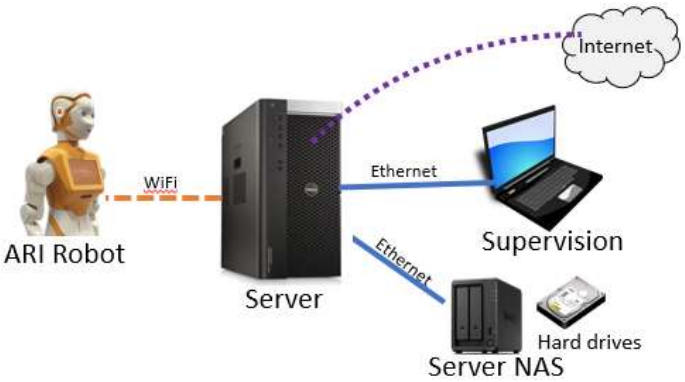


diagram of the communications used for wave 2

Wave 3: Autonomous experimentation

Following the feedback from the previous experimentation, each partner improved their algorithm. The functionalities of the robot remained the same, except that they are more precise. The most visible examples are the chatbot which has a new sentence or the robot move which is more efficient

Since the robot remains in the hospital, without the presence of the other partners, the supervision computer has been replaced by a control tablet. The experimenter now has control over the programme start-up and the choice to start the robot's dialogue with one or two people simultaneously.

For the moment, The records are simplified and are stored in the server.

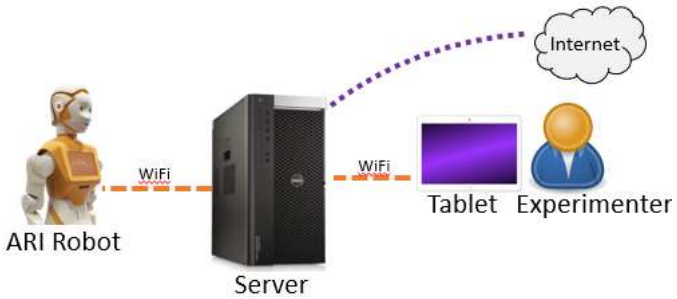


diagram of the communications used for wave 3

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

3. Description of the methodology and the results of three waves of intermediate evaluations

3.1 Summary of the assessments

This section describes the results of the interim evaluations conducted within the framework of the SPRING project between April 2022 and March 2023. The results described here concern three phases of evaluation, called respectively Wave 1, Wave 2, and Wave 3.

Table 1 presents the general characteristics of these assessments. The methodology and specific results for each wave are described in the following sections of the document.

Table 1. *General characteristics of the intermediary evaluations conducted within the SPRING project in relevant environments*

Study 1 (date)	Participants	Robot configuration	Data collected
Wave 1 (April 2022)	N=20 F= 19; M=1 Mean age = 74,9 y/o	- Wizard of Oz - Individual guided dialogues - No display on the screen - No displacement of the robot - Some arm movements of the robot were introduced -The robot's eyes moved randomly without really focusing on the person	Sociodemographics Acceptability E-Scale Robot features Perceived usefulness of potential use cases
Wave 2 (October 2022)	N=13 F=9 ; M=4 Mean age = 72,5 y/o	- No Wizard of Oz - Individual and multi-party autonomous interaction - Subtitles displayed on the screen - Orientation movement of the robot towards the user - No arms movements -The robot's gaze is always directed towards the person	Sociodemographics Acceptability E-Scale System Usability Scale Appreciation of robot features Ethics questionnaire Specific assessment of the robot's dialogues capacities
Wave 3 (March 2023)	N=20 F=16 ; M=4 Mean age =78 y/o	- No Wizard of Oz - Subtitles displayed on the screen - Individual and multi-party autonomous interaction - Orientation movement of the robot towards the user - No arms movements -The robot's gaze is always directed towards the person	Sociodemographics Acceptability E-Scale System Usability Scale Appreciation of robot features Ethics questionnaire

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

3.2 Specific description of each of the studies conducted in a simulated relevant environment (Study 1)

3.2.1 Study 1: Wave 1 experiment (Living Lab - Simulated environment)

Period: April 2022

Objectives

The main objective of the Wave 1 experiment was to evaluate the full initial architecture working on the ARI robot (version April 2022) with potential users of the solution (older adults using hospital services and their accompanying persons) in a simulated environment (i.e. in living lab conditions). This initial evaluation allowed the collection of audiovisual data and feedback from the potential users of the situation (older adults and accompanying persons), in contrast to other experiments conducted by the partner laboratories (carried out with young adults) that have been described elsewhere (deliverables D1.4, D2.1, D2.2, D2.4, D2.4, and D3.1).

Wave 1 assessments were structured to allow:

- User data collection during a task of human-robot interaction in a relevant environment to transmit to the technical partners responsible for the development and the programming of the SPRING system (vision, navigation, dialogue, emotional and engagement aspects).

- Gaining an understanding of the acceptability and usability of the SPRING robotic solution in a health facility context among potential users (older adults).

Participants

A total of 20 volunteers took part in this assessment. The volunteers were recruited through senior associations and former users of the hospital. After having obtained the agreement of the older adults to take part in the study, the researchers arranged a meeting with the participants for the test session in the living lab. The socio-demographic data corresponding to the participants are presented in Table 2.

Table 2. Socio-demographic data for participants in the "Wave 1" Living Lab assessment

Variables	Modalities	Total
Count	-	20 (100%)
Gender	Male	1 (5%)
	Female	19 (95%)

Age, years	*Min: 62* ; *Max*: 88 *Mean:* 74,9 *SD*: 7,07	
	[62 – 70[[71– 80[[81 – 88[	5 (25%) 11 (55%) 4 (20%)
Living Arrangements	Living alone Living with a partner	19 (95%) 1 (5%)

Material

The platform used for the tests was the robot ARI. ARI is a social assistant robot (PAL robotics) that is programmed, within the framework of the SPRING project, with a set of modules developed explicitly for this project. The robot mobilizes perceptual (audio and visual), proprioceptive (localization in space), conversational (language processing), and behavioral (adapting its proposals to the users' behavior) information. In the final stages of the project, the different modules will allow ARI to communicate automatically with several interlocutors simultaneously in a noisy environment. The robot will also adapt its behaviors according to what it perceives and understands from its interlocutors. This technical development work has been ongoing since the beginning of the project. The interactive versions of the modules are regularly integrated and evaluated in the different evaluation waves organized throughout the project.

The robot is programmed to provide support or distraction according to the user's request and the characteristics of the context of the human-robot interaction to act in an adapted manner. The general features of the ARI robot are described in Table 3.

Table 3. Technical presentation of the ARI robot

	**PHYSICAL CHARACTERISTICS** - Height: 1,68 m - Weight: 50 kg - Moves by rolling - Articulated arms (cannot carry anything) **INTERFACE** - Emergency stop button on the back - Touch pad - Animated eyes - Luminous ears **AUTONOMY** - 8 to 12 hours of battery life **CONNECTIVITY** - Wired and wireless, (Wifi, Bluetooth) **VISION** - 3 cameras (head, chest and back) **AUDIO** - 4 microphones	

The specific configuration of the robot for the Wave 1 evaluation comprised the following features:

- The use of the Wizard of Oz technique to handle the dialogues of the robot in real-time.
- Individual and multi-party guided dialogues, meaning that participants performed individual or multiparty verbal interactions with the robot (progressively more complex). The robot was provided with a female Text-to-Speech (TTS) voice (Acapela's UK English voice 'Rachel') set at 50% of the maximum volume). A series of tasks, related to reception and orientation for hospital users, was developed for this assessment.
- No dialogue subtitles were displayed on the robot's screen,
- The robot was located at a fixed point and no displacement of the robot was included,
- Some slight arm movements of the robot were introduced, but deactivated during the experiment
- The robot's eyes moved randomly, to increase its "lifelikeness", without really focusing on the person

Procedure

Participants were invited to enter a room where the Ari robot was located. The experimenter explained to the participants the use case "greeting and orientation" by inviting them to ask the robot questions normally asked when arriving at the hospital and needing an initial orientation at the facility. The robot was handled by the Wizard of Oz, to greet and introduce itself to the participants. Then, the participants could ask the robot different questions, for example, the

location of the consulting room, the cafeteria, or the toilets (Figure 5). Dialogues were handled by the Wizard of Oz to achieve individual or multi-party (collective) verbal interactions.

Figure 5. Example of interaction between the participant and the robot in the experiment Wave 1

An interaction of about 10 minutes followed, for which the robot was programmed with various dialogue contents that allowed it to answer the users' questions (see Table 4). In this first assessment, the robot's screen was not used to display any information or subtitles.

Table 4. Examples of dialogue contents managed by the robot using the Wizard of Oz technique

Type of question asked by the user	Robot responses
User 1: *"Can you help me find the toilets?"*	To User 1: *"For you, the toilets are opposite the dining room."*
User 2: *"Do you know if lunch is provided?"*	To User 2: *"And for you, lunch is provided for patients here at the clinic if you have consultations in the afternoon. For companions, there is a cafe on the ground floor."*
User 1: *"Can you tell me when the cafeteria opens?"*	To User 1: *"From 11 AM to 4 PM."*
User 2: *"ARI, could you tell me if my appointment is at 11 AM?"*	To User 2: *"Sorry, I don't have this information. You have to go to the reception."*
User 2: *"Can you tell me how to get to the reception?"*	To User 2: *"The reception is just behind me. Remember to take a ticket on your right."*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Assessments

At the end of the interaction task a comprehensive set of assessments was conducted to obtain a complete overview of the users' experience and their perspectives on the robot and its usefulness. These assessments included:

(A) Acceptability of the system was assessed using the acceptability questionnaire (AES). For each of the 6 items on the scale, the participant was invited to argue his response (scale in annex 1).

(B) Participants were then asked to answer two questions regarding their subjective experience of the robot:

- Rating the *appearance/design* of the robot (using a 5-point Likert Scale and qualitative feedback)
- Rating his/her *feeling about the interaction* with the robot (using a 5-point Likert Scale and qualitative feedback)

(C) Participants were invited to provide some examples of improvements that could be made to the robot, if relevant to the participant (using qualitative feedback)

(D) Finally, the perceived usefulness of different potential applications for the SPRING robot, that were presented to the participant, was examined. These use cases included:

a) The robot provides a **reception or welcoming** of the hospital users
b) The robot delivers information and reminders on **preventive hygiene measures**
c) The robot provides **assistance to patients throughout the care process**
d) The robot provides **orientation and guidanc**e related to the hospital services
e) The robot provides **entertainment activities** for persons who attend the hospital

Usability was not assessed in Wave 1 evaluations as the experiment was conducted through the use of a Wizard of Oz technique.

The person was invited to rate the perceived usefulness for each type of use case using a Likert scale (1: totally useless, 2: rather useless, 3: slightly useful, 4: useful, 5: totally useful), in the same way as for the acceptability scale, the participant was invited to argue his answers.

Results

(A) Acceptability of the robot - Wave 1

This section presents the results obtained in the evaluation using the AES (Acceptability Scale). Figure 6 displays the total scores given by the 20 participants on the AES.

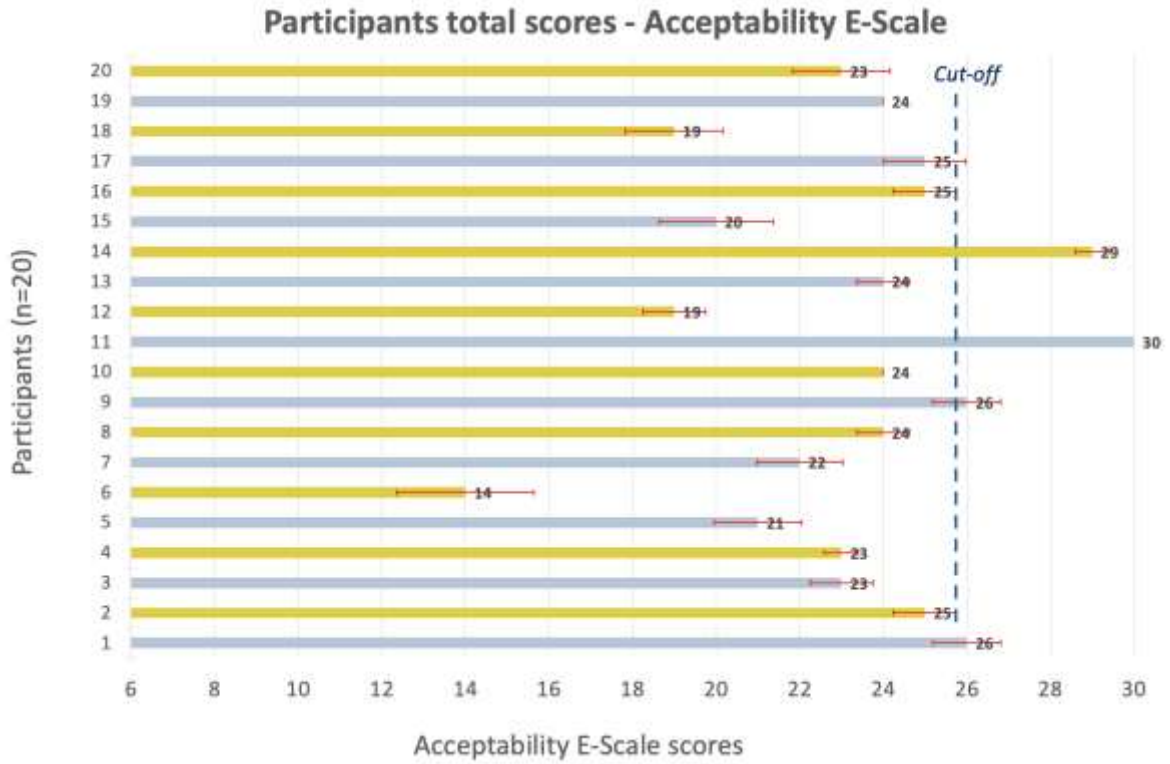


Figure 6. Global scores (AES) for the assessments in the Wave 1 (n=20 participants)

Participants in Wave 1 reported a mean total score of 23.3 (SD=0,76) over 30. This sample median was 24 over 30. Acceptability E-Scale cut-off is set to 25,8 over 30, indicating a satisfactory acceptability level, and four participants reported scores above the threshold. Wave 1 participants' scores ranged from 14 to 30 over 30. 15% of the participants reported a score lower than 20, 65% between 20 and 25,8, and 20% above 25,8. Figure 7 presents the mean scores on each dimension of the AES for the whole sample in Wave 1 (n=20 participants).

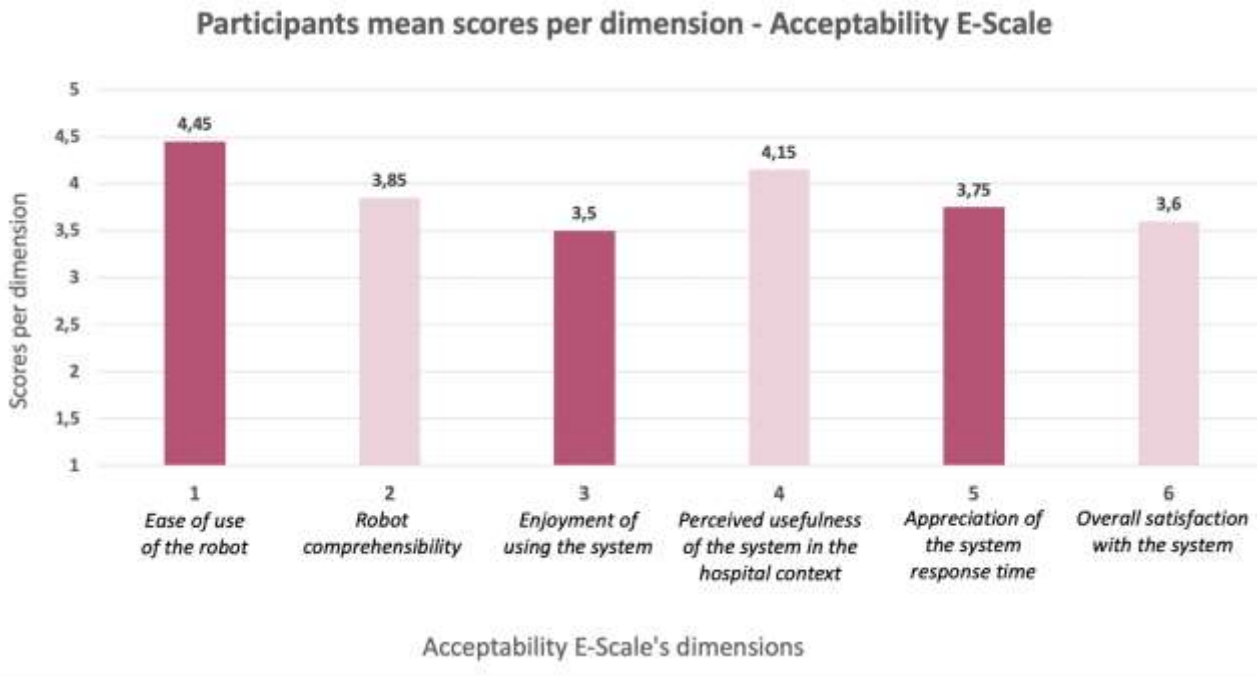


Figure 7. Mean acceptability scores (AES) for the assessments in Wave 1 (n=20 participants)

Participants' scores on the different dimensions of the Acceptability E-Scale are not homogenous with higher reported scores on dimension 1 - ease of use (m=4,45 ; SD=0,69) and dimension 4 - perceived usefulness (m=4,15 ; SD=0,88). Lower scores were reported for dimension 2 - Robot comprehensibility(m=3,85 ; SD=1,09), dimension 3 - enjoyment of using the system (m=3.5 ; SD=1.15), dimension 5 - Appreciation of the system response time (m=3,75 ; SD=0,91) and dimension 6 - overall satisfaction with the system (m=3.6 ; SD=0.94). On an interesting side, the lowest scoring dimensions are those with the highest standard deviations, meaning greater variations around the means. It can also be illustrated by examining the median scores (med=5 for dimension 1 ; med=4 for dimensions 2 to 6).

In the following section, we provide an analysis of the participants 'verbatim associated with the questions about the acceptability of the robot.

Ease of use of the robot (AES)

Most participants mentioned that the robot was easy to use, basically because its main role was to provide information, and that no particular manipulation of the robot was requested:

> *"It [the robot] provides an answer to the questions asked"*;
> *"It understands the questions well enough"*;
> *"The answers are clear"*;
> *"The robot answered my questions even though I used long sentences. It was able to grasp the essence of my questions".*

However, some participants mentioned some limitations regarding the usability of the robot:

- The use of the robot would require assistance: *"The robot is rather easy to use if I am accompanied"*.
- Some participants encountered some difficulties to understand the robot's answers: *"I had some difficulties deciphering the answers of the robot"*
- The robot is easy to use in a situation known/prepared in advance: *" It is easy because I know I am talking to a robot and the situation is known in advance, I am not surprised"*;
- The participants need to use it in a real situation to judge if it is really easy to use: "*I did not have to use it in a real situation"*

Robot comprehensibility (AES)

Some participants said that the robot was rather easy to understand:

> *"The robot was pretty easy to understand if the answers (dialogues) are not too complex*".

However other participants mentioned some limitations regarding robot comprehensibility:

- The answers of the robot were too fast: *"The succession of information to tell how to go to a specific place is too fast to remember the robot's answer"*
- The sound/the voice of the robot was too weak/low, especially if the user has hearing problems: "*As a hearing-impaired person, I did not always understand some of the answers well";*
- The language of the robot was too robotic/mechanic: *"the phrasing is sometimes too robotic", "it seems to react only to certain keywords"; " I would like more variation in the voice, it is a little too automatic"*

Enjoyment in using the system (AES)

Some participants gave a variety of reasons to explain why they had enjoyed the interaction with the robot:

- The robot is perceived as an innovative technology: *"Robots are the future"*
- The robot is fun: *"I enjoyed the session with the robot that I took as a game"*
- The robot is reassuring: *"The robot is reassuring; I was not intimidated"*
- The robot is always ready to help: *"The robot is always available to answer to people"*

On the contrary, several participants noted that the technical problems encountered during the interaction with the robot negatively impacted the excitement and fun initially associated with their encounter with the robot:

- The robot is annoying due to technological limitations: *"The robot is fun at least at first, but then it quickly becomes annoying due to its limitations"*
- The voice interaction is more pleasant than mobility and physical behavior: *"I enjoyed more the interaction than the navigation and arms gestures"*
- The robot could reduce human contact: *"I enjoyed the robot as a game, but I miss the human contact that remains the warmest, even irreplaceable"*

Usefulness of the robot in the hospital context (AES)

The participants perceived the following benefits of the robot when used in a hospital context:

- Streamline the patient journey: *"the advantage of the robot is to streamline lines at the information desk"*
- The robot can provide occupation and entertainment during hospital waiting times: *"[the robot] can help to "humanize" the waiting time, so that people waiting for consultations do not get bored"*
- The robot can be useful to provide information and orientation to hospital users: *"At the front desk or in the hallways, the robot can provide information to lost persons"*
- The robot can help to offload the work of hospital staff "*[the robot] could be useful perhaps to offload the work of reception staff" ; "the robot is a complementary support to the staff"*
- The robot can compensate for the lack of personnel in the current critical context of shortage of care professionals: *"[the robot] is useful because of the situation of lack of nursing personnel"*
- The robot behavior is always in a stable/good mood towards the people consulting the day care hospital: *"the robot raises the question of the 'non-human' in the hospital, at the same time, it spares you its mood variations and stress, which is not the case for the human workers"*

However, other participants perceived some disadvantages related to the use of robots in a hospital context:

- The use of the robot induces a risk of dehumanization: *"The robot is moderately useful, the medical and paramedical staff and all the persons in the hospital are psychologically indispensable and reassuring for me"*
- The robot must remain a complement to human accompaniment: *"[the robot] is useful, but it must remain a complement to real persons". "The lack of humanity [or the robot] is to be considered in a hospital environment".*
- The robot can only be useful if the technology is advanced enough for fluid human-robot interactions: *"It all depends on the reactivity and the length of the questions and answers. If the interaction is not fluid, there is a loss of time and therefore ultimately a need to consult a person".*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

- This type of robotic solution will not be suitable for older adults who frequent the hospital: *"Many people over 75 may be confused [with the robot] because they are not used to anything modern"*
- The robot could be replaced by other simpler technologies, considering that at the time of the evaluation, the robot visual dimension is not sufficiently exploited.: *"Maybe something can be done with the robot [at the future], but at the moment a simple tablet/screen providing information could play the role"*
- The presence of the robot underlines a situation that is already critical enough in the context of health care institutions and that would represent a degraded solution: *"The robot is palliative for a lack of personnel"*

Appreciation of the robot's response time (AES)

Some participants thought that the answers of the robot were quite quick *"The answers were quite immediate"*, but others found that the robot's answers were slow: *"Some of its responses were slow, but a robot is expected to respond slowly"; "It is acceptable but a little slow"*

Overall satisfaction with the robot (AES)

Some participants were very positive regarding the experience with the robot:

"This is an interesting project"
"The robot is not finished but it is already amazing".

Other participants felt that the experience was not entirely positive for a variety of reasons:

- The robot, for the time being, is not sufficiently capable of responding to the user's demands: *"I found it not responsive enough"*
- The interaction with the robot does not provide any particular entertainment, the interaction experience is not very enjoyable: "The interaction was not fun enough"
- The robot's mobility capabilities are not sufficiently exploited at this time: *"I would have liked to see him move [but it doesn't]"*
- The information content currently present in the robot is not sufficiently rich.: *"it [the robot] always repeats the same information"*

B. Subjective experience of the robot - Wave 1

Participants were also asked to rate the appearance (aspect, design) of the robot using a Likert scale *(1= Very unpleasant, 2=Unpleasant, 3=Neither pleasant nor unpleasant, 4=Pleasant, 5=Very pleasant)*. The mean score for the 20 participants was 3,6 (SD = 1,10; median=4) (See Figure 8). Overall, the robot's appearance is rather appreciated. However, we also find some opposite views with respect to the actual design of the robot.

Participants were asked to rate from 1 to 5 their feelings during the interaction with the robot using a Likert scale *(1= Very negative, 2=Negative, 3=Neither negative nor positive, 4=Positive, 5=Very positive)*. The mean score for the 20 participants was 4 (SD = 0,73, med=4) (See Figure 8). Overall, the results illustrate a diversity of positions with respect to the overall experience of interaction with the robot, with both positive and negative attitudes toward the robot.

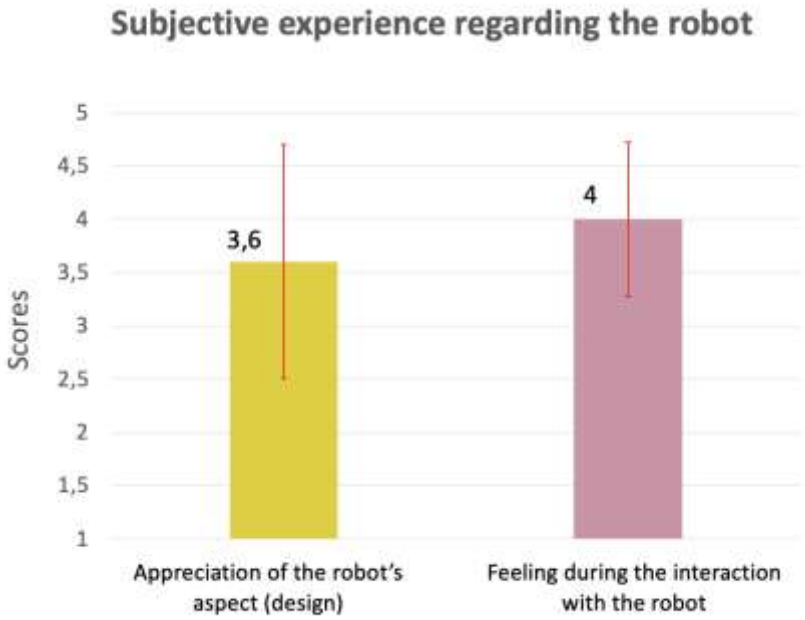


Figure 8. Mean scores of the assessment regarding the subjective experience during the interaction with the robot - Wave 1 (n=20 participants)

With respect to the appearance of the Ari robot, some qualitative feedback was globally positive. Participants declared being satisfied with:

- The design of the robot: "*The robot is rather pleasant because of the roundness of its appearance. The moving eyes are a plus; the robot has a nice look*"; "*The robot is very nice thanks to its mermaid skirt*"; "The colors are neutral and discreet"
- The size of the robot: *"The size of the robot is adequate"*
- The general feeling of the design: *"It's not scary"*

However some participants expressed some limits about the robot's design:

- The design, the aesthetic: *"I don't like its design, its base is very heavy", "It is a bit monolithic"; "I would like it to be more humanoid, with legs, more aesthetic"*
- The size: *"It is too big"*
- The colors: *"Its color is not a color. It would be more beautiful with different colors, sharp, bright, happy"*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

With regard to the feelings of the participants during the interaction with the robot, the answers obtained in the qualitative feedback were varied. Some participants expressed positive feelings during the interaction with the robot:

- The feeling of experiencing real innovation: *"The robot is progress"*
- An enjoyable experience largely linked to the robot's eye-catching appearance: *"The robot is very pleasant to look at"*
- The experience of interacting with the robot did not involve any kind of fear, apprehension or a feeling of weirdness.: *"As we were prepared for the tests of the robot there was no apprehension"*
- The interaction with the robot was easy, simple and accessible: *"The robot is user-friendly"*; *"There was no problem, the interaction with the robot was natural"*

On the contrary, other participants described a rather mitigated experience, noting some aspects of displeasure with respect to the interaction with the robot:

- Have the impression of a very mechanical and controlled interaction: *"We feel like we are in front of a machine"*
- Experience a feeling of annoyance because of technological limitations: *"I was disappointed because it did not take me [provide an accurate information] where I wanted to go"*
- Declaring some doubts about the usefulness of the robot: *"I was happy to do the experiment but I'm not completely convinced of its usefulness"*

C. Suggestions for the improvement of the robot or the robot-related services - Wave 1

The participants provided some suggestions to improve the interaction with the robot.

- The robot could make a small presentation of what it can do (type of information possibly given by the robot) would be useful. What is it used for?
- The robot could be more friendly: "*It could provide friendly greetings with explanations of the type of questions to ask to be included".*
- The robot could be more responsive: "*The robot is not very responsive, it should behave in a way that is closer to that of a living being".*
- The robot should be more friendly, *"it should include a smile" "The robot could have a mouth"*
- The robot could be personalized: *"put a suit on it"*
- The robot's voice could be adapted to the user: *"Sometimes [the voice is] difficult to understand, "the volume is too low", "it should be possible to adapt the volume of the voice to the user"*; *"The diction of the robot should be improved"*

D. Perceived usefulness of potential use-cases for the SPRING robot - Wave 1

The participants' opinion regarding potential use cases for the SPRING robot in the hospital context was assessed using a 5-point Likert scale *(1= Useless, 2= Rather useless, 3= Neither useless nor useful, 4= Rather useful, 5= Useful)*, followed by qualitative feedback.

The participants considered the most relevant use case for the robot in the hospital context (1) the provision of orientation and guidance of users regarding the hospital services, followed by (2) the provision of information regarding preventive hygiene measures, (3) the reception and welcoming of hospital users, (4) the offer of entertainment activities, and lastly (5) the provision of assistance to patients throughout the care process (See Figure 9).

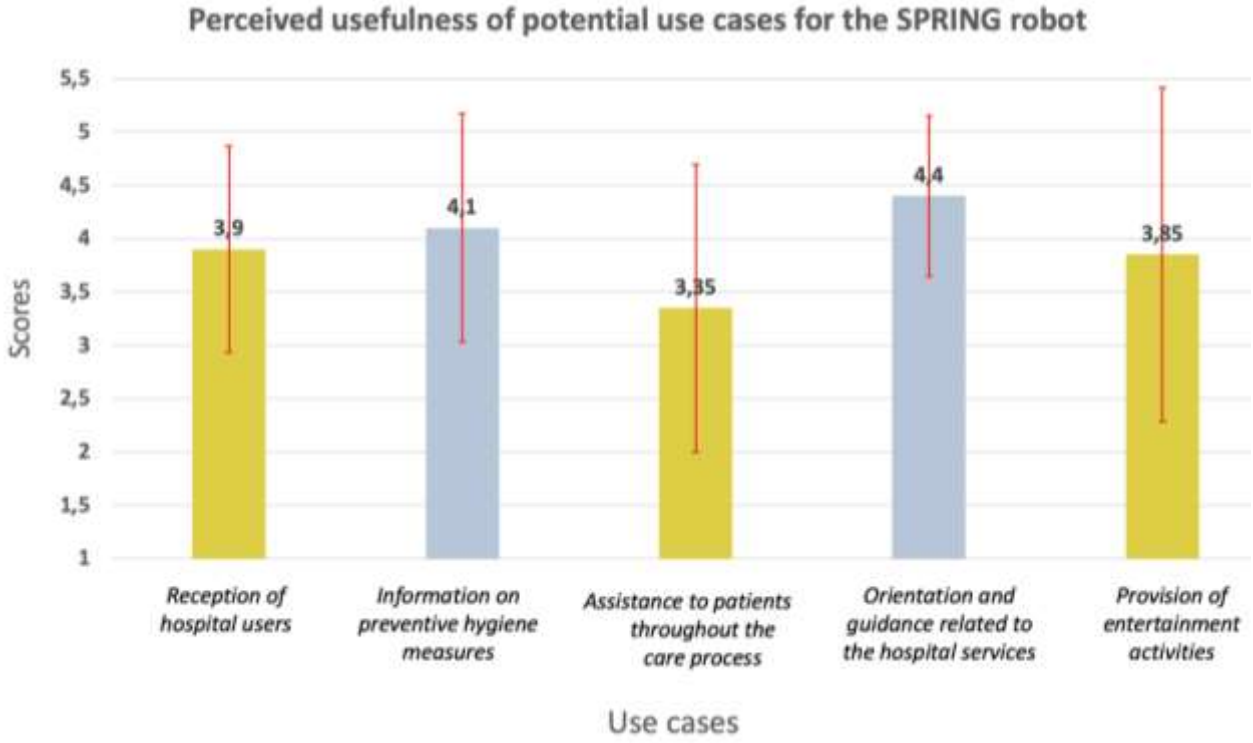


Figure 9. Perceived usefulness of potential use-cases for the SPRING robot in the hospital context

Qualitative feedback regarding the perceived usefulness of potential use cases for the SPRING robot in the context of a hospital is presented in Table 5.

Table 5. Qualitative feedback regarding perceived usefulness of potential use cases for the SPRING robot in the context of a hospital

Use cases	Positive points	Limits

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

(1) Reception of patients and their carers	This case was globally perceived as useful, particularly in case of staff shortage	- The robot-based service should be enjoyable to use - The verbal information provided by the robot needs to be short and not have too much information at once. - A familiarization with the robot is required before taking advantage of it: "*You have to get used to the interaction with the robot first*" Some participants wonder if this service will be adapted to the psychosocial needs of patients: "*Will it be able to reassure a worried patient?*" ; -There may be some limits to the accessibility of the robot for certain publics. *"Will it be adapted even for disabled people?"*
(2) Delivering information and reminders on preventive hygiene measures	- This use case could provide a complement to the role of the hospital professionals *"It could be a reminder if you didn't wash your hands or didn't wear the mask properly".* - A positive point of the robot in promoting preventive hygiene measures would be that the robot is not sensitive to verbal aggression: *"Hospital users can therefore express their dissatisfaction to the robot without repercussions".*	- The robot's screen could be used more effectively. Indeed, several participants thought that the interactive screen could be used to show videos on the sanitary rules to be respected. *"Very useful with a short video reminding you how to wash your hands*

(3) Providing assistance to patients throughout the care process	-The use case is considered as useful if the help concerns paperwork to do before a consultation -The hearing impaired and disoriented elderly need repetition, and robot is not bothered to repeat things several times	- The use case can be unnecessary if the same service is provided by hospital staff at the reception - The quality of human service is considered superior to that offered by a robot: "*a human being will do this better than a robot*" The robot can neither understand the needs of all users nor adapt to them. "*will it [the robot] understand if the patient doesn't seem to understand*" -Some concerns regarding data collection by the robot were mentioned: *"the robot shouldn't record users' data"* - The use case could be improved if the screen of the robot is used to show videos in addition to voice
(4) Providing orientation and guidance to hospital users	- The use case was considered useful to help users get easily to the different places in the hospital *"Very useful as a complement to building signage*" - The features of the robot (screen, mobility) could help to improve the quality of the service provided to users, but these applications should be used. "*The robot can put a map of the hospital and show the way on the screen*", "*It can accompany the patient to the requested place*" *"It could be useful if the robot really accompanies the people"*	- The quality of human service is considered superior to that offered by a robot: "*Contact with humans is preferable*"

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

(5) Entertainment	The use case is considered useful since it could improve waiting time and ease some tension between visitors *"The robot could help pass the time"* - This use case could help the patients to relax (without disturbing other patients).	- There may be a risk of tension between hospital users if several people want to use the robot at the same time *"It could be a source of disputes"* - For people wishing to be at calm, they could be disturbed by the robot - The service could be improved by offering some music *"It could play songs"*

3.2.2 Study 1: Wave 2 experiment (Living Lab - Simulated environment)

Period: October 2022

Objectives

The main objective of the Wave 2 experiments was to evaluate an updated version of the architecture working on the ARI robot (version October 2022) with potential users of the solution (older adults using hospital services and their accompanying persons) in a simulated hospital environment (i.e. in living lab conditions). Wave 2 assessments were structured to:

- Continue the process of user data collection and its provision to the technical partners in charge of the development and improvement of the technical modules of the robot (vision, navigation, dialogue, emotional and engagement aspects).

- Conduct further collection of user feedback regarding acceptability, usability, and satisfaction with the SPRING robotic solution.

- Gather specific user feedback regarding the dialogue capacities of the SPRING robot (in collaboration with the partner HWU)

- Begin to explore the opinions and perspectives of potential users regarding the ethical challenges related to the use of socially assistive robots in a healthcare delivery setting.

Participants

A total of 13 volunteers took part in this assessment. The volunteers were recruited through senior associations and former users of the hospital. After having obtained the agreement of the older adults to take part in the study, the researchers arranged a meeting with the participants for the test session in the living lab. The socio-demographic data corresponding to the participants are presented in Table 6.

Table 6. Socio-demographic data for participants in the "Wave 2" - Living Lab assessments

Variables	Modalities	Total
Count	-	13 (100%)
Gender	Male	4 (31%)
	Female	9 (69%)
Age, years	*Min:* 64 ; *Max*: 81 Mean: 72,5; SD: 4,53	

	[64 – 69[[70-75[[76 – 81[	3 (23%) 7 (54%) 3 (23%)
Living Arrangements	Living alone Living with a partner	10 (77%) 03 (23%)
Education Level	High School Diploma College Degree	03 (23%) 10 (77%)

Material

The robot ARI (PAL robotics) configured using the different modules developed for the SPRING project was used as a platform for the assessments. A global description of the ARI robot and the specific modules developed for this platform in the SPRING project have been described in Section 3.2.1 (Material) of this report.

The specific configuration of the robot for the Wave 2 assessment comprised the following features:

- Dialogues were handled automatically by the robot in real time. (no Wizard of Oz)
- Individual and multi-party interactions were proposed, meaning that participants performed individual or collective (two persons) verbal interactions with the robot. The robot was provided with a female Text-to-Speech (TTS) voice (Acapela's UK English voice 'Rachel') set at 50% of the maximum volume). A series of tasks, related to reception and orientation for hospital users, was developed for this assessment.
- Dialogue subtitles were displayed on the robot's screen,
- The robot moved towards the user at the beginning of the interaction.
- Not arm movements of the robot were introduced
- The robot's gaze was always directed towards the person

Procedure

Participants were invited to enter a room where the Ari robot was located. The experimenter explained to the participants the use case ("greeting and orientation") by inviting them to ask questions directly to the robot related to their arrival in the hospital. The robot was programmed to greet and introduce itself to the participants automatically. Then, the participants could ask the robot different questions, for example, ask where the consulting room, the cafeteria, and the toilets were located.

A total of ten spoken interactions between the robot and volunteer participants were held. Five were multi-party conversations, in which two participants interacted with each other and the robot; five involved a single participant conversing with the robot (See Figure 10). Two of the multi-party interactions were 'open' dialogues i.e. participants were free to choose their own enquiries.

One was fully task-based, which meant the participants were supplied with a set of 6 consecutive tasks that had been designed to illustrate the key functionality of the system. The individual tasks were represented as an image (see Appendix A for an example task sheet) to avoid prompting participants to use particular phrases when talking to the robot. The remaining two MPC involved a combination of first open and then task-based dialogue.

Figure 10. Example of interaction between the participant and the robot in the experiment Wave 2

An interaction of about 10 minutes followed for which the robot was programmed with various dialogue contents that allowed it to answer the users' questions, either in the individual or in the multi-party dialogue condition (see Table 7).

Table 7. Examples of dialogue contents managed by the robot - Wave 2

Type of question asked by the user	Robot responses
User 1: *What paper do I need to register at reception?*	To User 1: *"You need your convocation and your social security card. The secretaries will explain everything to you. They will help you if you forget"*
User 1: *"I have just arrived, what should I do?"*	To User 1: *"You must take a ticket, you will be called to register"*
User 1: *"Can my companion accompany me during the consultations?"*	To User 1: *"You can ask the secretaries at the entrance or the nurses"*
User 1: *"Can you point me to the bathroom?"*	To User 1: *"To find the nearest bathroom, leave the dining room and go right towards the door with two round windows. The bathroom will then be on your left"*
User 1: *"can you point me to the lift please?"*	To User 1: *"To get to the lift, leave the dining room, turn right and go through the doors with two round windows. Continue straight down the corridor*

	through further doors until you reach the lift landing."

Assessments

At the end of the interaction task a comprehensive set of assessments was conducted to obtain a complete overview of the users' experience and their perspectives on the robot and its usefulness. These assessments included:

(A) Acceptability of the system was assessed using the acceptability questionnaire (AES) was administered. For each of the 6 items on the scale, the participant was invited to argue his response (scale in Annex 2).

(B) Participants were then asked to answer two questions regarding their subjective experience of the robot:

- Rating the *appearance/design* of the robot (using a 5-point Likert Scale and qualitative feedback)
- Rating his/her *feeling about the interaction* with the robot (using a 5-point Likert Scale and qualitative feedback)

(C) Participants were invited to provide some examples of improvements that could be made to the robot, if relevant to the participant (using qualitative feedback)

(D) Usability of the system was assessed using the System Usability Scale and qualitative feedback was requested from participants to explain the notes attributed in the different dimensions of the scale (see Annex 1).

(E) A specific assessment was done in this wave regarding the appreciation of the dialogue capabilities of the robot (HWU questionnaire). The questionnaire is composed of 6 items that are rated using a 5-point Likert Scale and qualitative feedback (see Annex 6).

(F) Finally, a semi-structured ethical interview with the participants was conducted (see Annex 4).

Results

(A) Acceptability of the robot - Wave 2

This section presents the results obtained in the evaluation using the AES (Acceptability Scale).

Participants in Wave 2 reported a mean total score of 21.7 (SD=0.81) over 30. This sample median was 20 over 30. Acceptability E-Scale cut-off is set to 25.8 over 30, indicating a satisfactory acceptability level, and two participants reported scores above the threshold. Wave 2 participants'

scores ranged from 19 to 28 over 30. 54% of the participants reported a score lower or equal to 20, 30% between 20 and 25.8, and 16% above 25.8.

Figure 11 presents the total scores on the acceptability scale for the 13 participants who took part in the Wave 2 assessment.

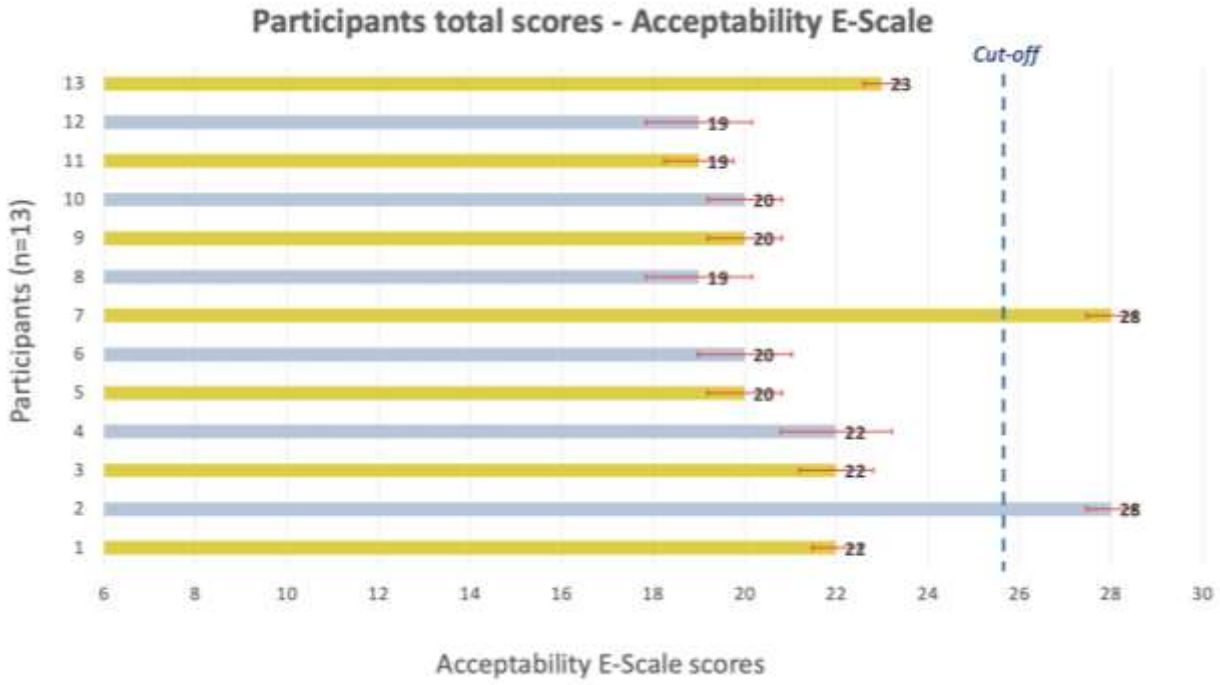


Figure 11. Global scores (AES) for the assessments in the Wave 2 (n=13 participants)

Participants' scores on the different dimensions of the Acceptability E-Scale are more homogenous in Wave 2 - with a median of 4 across the 6 dimensions - than in Wave 1 even though the reported scores failed to reach a satisfactory acceptability level (25,8). In Wave 2 the highest scoring dimensions are dimension 4 - perceived usefulness (m=4 ; ET=0,82 ; med=4) and dimension 5 - Appreciation of the system response time (m=4 ; SD=0,71 ; med=4). Lower scores were reported for dimension 1 - ease of use of the robot (m=3,5 ; SD=0,97 ; med=3), dimension 2 - robot comprehensibility (m=3,7 ; SD=0,85 ; med=4), dimension 3 - enjoyment of using the system (m=3 ; SD=1,12 ; med=3) and dimension 6 - overall satisfaction with the system (m=3,5 ; SD=0,66 ; med=3).

Figure 12 presents the mean scores for each dimension of the acceptability scale for the whole sample (n=13 participants) who took part in the Wave 2 assessment.

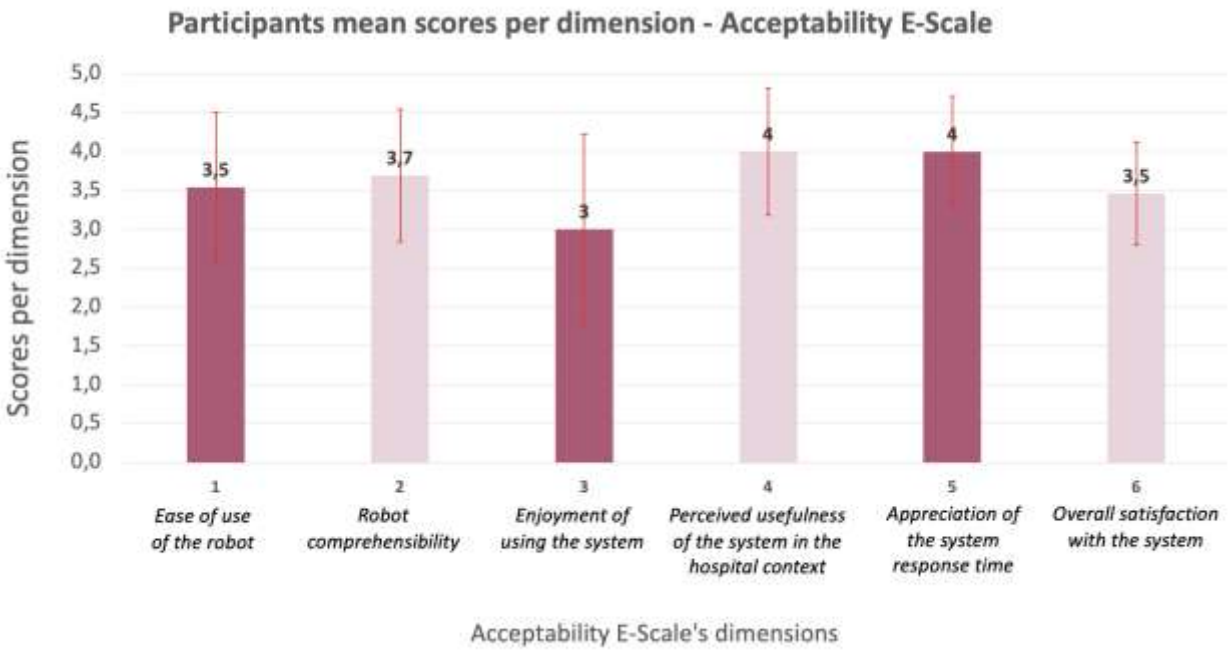


Figure 12. Mean acceptability scores (AES) for the assessments in the Wave 2 (n=13 participants)

Below we present the qualitative returns that the participants made with respect to each of the dimensions of the Acceptability E-Scale in the Wave 2 evaluation.

Easy of use of the robot (AES)

The robot was globally considered to be easy to use: "*You just have to wait to ask questions, to wait for the answer, so there is no difficulty of use".*

However, some aspects that limited the robot's ease of use were identified as well, for instance some participants found the robot difficult to use because they didn't understand how to ask the questions to get an appropriate answer from the robot:

> *"I couldn't figure out how I should ask its questions. It was annoying".*

Others emphasized the robot's comprehension problem:

> "*You have to speak loudly and slowly enough so that it will understand. When you see what is displayed in relation to what you said, often it is gibberish and it (the robot] cannot decipher what is being said".*

The ease of use depends on the questions asked to the robot:

> "*Easy questions were easily answered by the robot but when the question is a bit complex. The robot was stuck".*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Some persons mentioned that the robot might seem difficult to use to people who are not familiar with technology:

> *"I am comfortable with the robot because I am also comfortable with computers, so maybe that's why it's easy to use for me, it won't be the general feeling for older people who don't know".*

Robot comprehensibility (AES)

Globally, participants observed that the robot was rather easy to understand:

> *"We understood very well what the robot said*".

The fact that the words were written down on the screen (subtitles) facilitated the comprehension of the robot's dialogues for some users:

> *"The answers and questions are written down, which makes it easier to understand"*

> *"It [what the robot says] is written on the screen and that is good"*

However, some aspects that limited the robot's comprehensibility were identified as well, for instance the robot's answers were not always appropriate to the questions:

> *"To say where the restrooms are or the elevator, that's fine. To say where the services are, for example, the different medical services, it was very complicated. The answer was completely wrong".*

> *"When I asked him to go somewhere or to meet someone, he answered me by telling me about a meal".*

Enjoyment of using the system (AES)

Participants had mitigated opinions regarding the system's agreeableness, although they understood that the work was not finished but still in progress. It was admitted it was normal for the robot to be gradually improved:

> *"If you ask [the robot] where the toilet is, it's fine. The answer is also good if you ask where the elevators are. But to find the nurse or the doctor, it's a bit confusing for the robot, but it can still tell you to wait. That's understandable, right? And when the program is perfectly finished, the principle is good"*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

The participants were mostly disappointed and frustrated by the interaction with the robot when they did not get relevant answers to their questions*:*

> *"People would like to have information about the place where they can find a doctor if they come for a consultation or a patient if they come for a visit, or a service (radiology, sampling) whereas the information about the cafeteria or the toilets given by the robot is of secondary interest".*
>
> *"I had the impression that the robot did not understand my questions".*
>
> *"I did not have a real conversation with the robot because the answers had no link with the questions".*
>
> *"The same answer is given over and over again about food and meals, it's very frustrating"*

Another reason that limited the pleasantness of the experience was the comparison with human service: *"a person is more pleasant".*

Usefulness of the robot (AES)

Participants mentioned different possible useful use cases for the robot in the context of a hospital

- To relieve the reception staff: *"the robot could relieve the reception staff"*
- To help the nursing staff at night in the hospital: *"I will see the robot useful for the night in the hospital"*
- To make up for a lack of personnel: *"If there is a strike in the hospital. It will be very useful. If there is a shortage of staff"; "If I was in a normal hospital visiting situation, I would be happy to have a robot if I can't find someone else"*.
- To free up staff time for other tasks: "*Given the lack of staff, I think it would allow people to be deployed on other tasks".*
- To make the hospital more welcoming for visitors: "*I find a robot at the reception desk rather nice for people who come for consultations or to visit patients"*
- To entertain the hospitalized: "i*t would be useful for the sick persons in the hospital, they would know it. They would go and ask for it without any problem. I would even be an amusement for some of them. It will be important to communicate about the robot for the hospitalized patients"*

Appreciation of the robot's response time (AES)

The average response time of the robot was considered satisfactory by some participants: *"It's fine. These consultations are more for the elderly, so they need time to get used to the screen, so there is always some apprehension".*

However, some limits on the robot's response time were pointed out:

- The waiting time is variable and may generate uncertainty: *"The answer is very interactive, there are other places where there is really a waiting time, the person will wonder if he has understood correctly"*
- The waiting period is sometimes too short: "*Sometimes they don't wait for your entire question before they start talking"*.

Overall satisfaction with the robot (AES)

Participants described the overall satisfaction of the robot in a mitigated manner. Despite the recognized potential in the use of this type of robot, in the context of a hospital, the technology is not yet considered finalized, there are many bugs and technical problems at the moment. At this stage of the project development, the robot is not considered to meet the expectations of the users.

> *"[My overall satisfaction] is ambivalent"*
>
> *"The robot has a good responsiveness but it is not adapted to the questions, it has a problem of understanding"*
>
> *"Very little satisfaction. This robot is rather intended to provide information on very practical services, may for example, on consultations, it does not know eh, so..."*
>
> *"There are still many things to finish. When it will be finished, we will be at the last judgment to see if everything goes well at that moment".*
>
> *"Compared to the expectations I have [of the robot] and what it delivers today? I am disappointed, I am sorry"*

B. Subjective experience of the robot - Wave 2

Participants were also to rate the <u>appearance (aspect, design) of the robot</u> using a Likert-scale *(1= Very unpleasant, 2=Unpleasant, 3=Neither pleasant nor unpleasant, 4=Pleasant, 5=Very pleasant)*. Mean score for the 13 participants was 3,85 (SD = 0,90; median=4) (See Figure 13). Overall, the

robot's appearance was rather appreciated, although some points of disagreement were also noted.

Participants were asked to rate from 1 to 5 their feelings during the interaction with the robot using a Likert-scale *(1= Very negative, 2=Negative, 3=Neither negative nor positive, 4=Positive, 5=Very positive)*. Mean score for the 13 participants was 3,62 (SD = 0,96, median= 4) (See Figure 13). Overall, the results illustrate a diversity of positions with respect to the overall experience of interaction with the robot, with both positive and negative attitudes toward the robot.

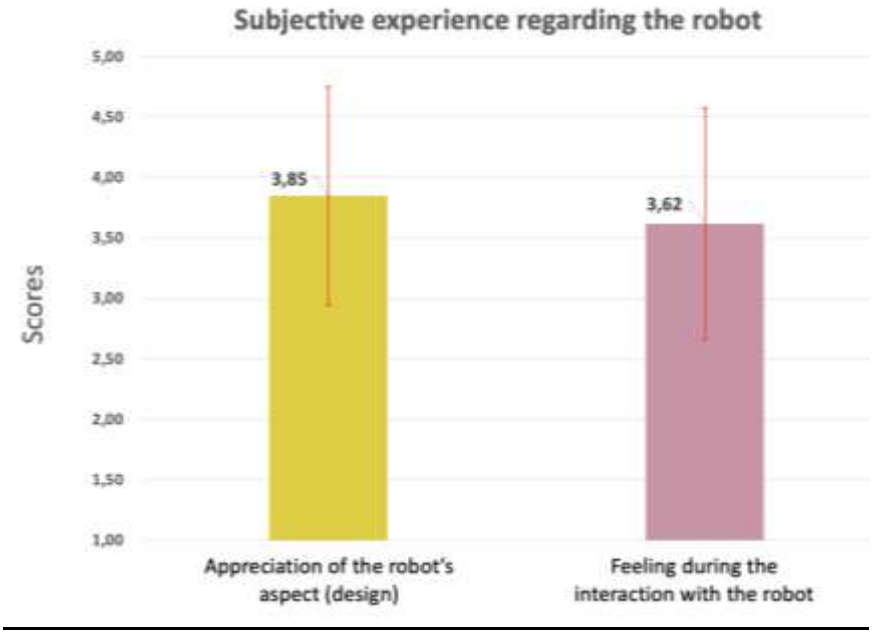


Figure 13. Mean scores of the assessment regarding the subjective experience during the interaction with the robot - Wave 2 (n=13 participants)

With respect to the appearance of the Ari robot, some qualitative feedback was globally positive. Participants declared being satisfied with the design of the robot:

- Pleasant: "*I liked it, I think it has a good head. There are eyes that move, so we also have the impression that it follows the person. Then the roundness, so it gives, it gives a bit of warmth. with the shape too. It's important too*".
- Nice: "*It looks young and slender. We can say it is nice. The colors are not bad. The head with the eyes of the small balls is good, it's nice. We immediately know it is a robot*"
- The current design seems balanced and sufficient: *"The design is just acceptable"; "The design is neutral".*

However, participants also mentioned some critical remarks with respect to the design of the robot:

- The design is too much of a machine: "*The form is not very pleasant, it is really machine, it is synthetic, it can be improved, that it is more human"*
- The design is too humanoid: *"I don't like robots that have a humanoid appearance. I would like it to be more like a coffee machine"*
- Some persons did not like the tone of the voice: *"The tone of voice is always the same so it is annoying"*

Regarding the feelings of the participants during the interaction with the robot, the answers obtained in the qualitative feedback were varied. Some participants expressed positive feelings during the interaction with the robot:

> *"Me, I'm comfortable [with the robot]. It's okay, I'm also comfortable with computers so maybe that's it..it's not going to be the general feeling."*
>
> *"It was an experience I enjoyed"*

Other participants highlighted several limitations in the experience of interacting with the robot:

> *"It's still very synthetic, yes, very digital..."*
>
> *"I didn't have a feeling at all, not at all"*
>
> *"I was disappointed, well, you can see the limits [of the robot]"*
>
> *"I thought it [the robot] would perform better for people."*

C. Suggestions for the improvement of the robot or the robot-related services - Wave 2

Some participants suggested ideas for improvement of the robot or implementation in the hospital:

- Provide explanations on how to use and limitations: "*Explain well what they can do: provide a small instruction manual: give the instructions for use, clarify their scope of work because currently they will not be able to answer some questions".*
- Position the robot so that it is immediately visible at the reception desk.
- Improve the interaction with the robot:" *the adequacy between the question asked and the answer that provides us that you think it should be improved" ; "it is in relation to the vocabulary, that it must be improved"*
- Improve the mobility of the robot: *"Making the arms move so that it can accompany the robot's speech, it can bring a bit of a bond."*

- In the context of a hospital it could be useful to have different robots with specialized functions:

  "Big hospitals, but basically there is always a receptionist that gives you all the information. I think it would be nice to have several robots and to specialize them: distraction robots, location information robots, consultation robot, doctor, care robot"

- The font used for the subtitles on the screen should be enlarged for the visually impaired: *"The subtitles should be a little larger for people with visual problems"*

- The robot's design could be a little more vibrant: *"We can use brighter colors. Maybe it will flash people better because it's a little neutral"* .

D. Usability of the robot - Wave 2

With respect to the evaluation of the usability of the device, the results of the evaluation made with the SUS scale showed that

In wave 2, the participants reported a mean score of 62 on the SUS scale (ET=15,45 ; med=65) which would be considered below average but still "OK" according to the SUS rating scale (refer to Figure 4, P5). As a reminder, a SUS score above 68 would be considered above average and anything below 68 is below average. Participants 5 and 12 can be considered as outliers and not taking their scores into account would significantly change the mean score (m=67 ; ET=11,08 ; med=68) making the mean score around the above average rate.

Figure 14 presents the scores given by the 13 participants on the SUS usability scale.

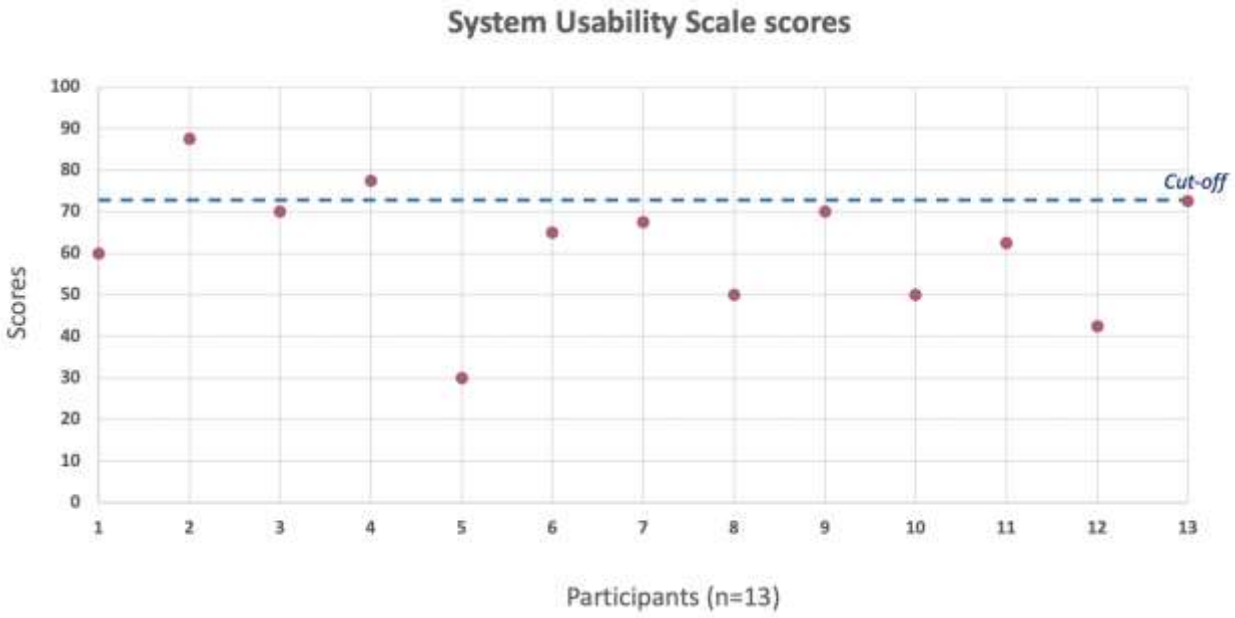


Figure 14. Scores for the scale of usability (SUS) for the assessments in Wave 2 (n=13 participants)

Qualitative feedback given by the participants regarding the usability assessment is presented below.

Interest in using the robot regularly in the hospital (SUS)

Participants globally agreed on the potential usefulness of the robot. However, they emphasized that they would prefer to have interactions with humans rather than with a robot, if the choice is given.

> *"I wouldn't use it all the time, but from time to time"*

> *"I think it has to be gradual [the introduction of robots in the hospital], right. You probably have to do a communication campaign beforehand. To reassure people…"*

Still, one volunteer noted that he would go to the robot first, but if he was not satisfied with its response, he might go to a human later.

> "*It depends, do I have a choice between going to a person? If I only have the robot and I must do it, I can't do it any other way. If there is a person in addition, I would like to use the robot, but if I am not satisfied, I'd always want to be able to have a person who can assist me, if I have not understood well* "

Perceived complexity of dialogues with the robot (SUS)

Concerning the question about the conversations with the robot being unnecessarily complex, some participants felt that there was no dialogue with the robot, which bothered them. Indeed,

they would have liked to have a real conversation with the robot, so the robot would have had to remember the information said earlier in the conversation.

> "*It's not a dialogue, I don't feel like I'm having a dialogue*"

There would be an issue of confidentiality when interacting with the robot in a public context which means that its answer can be heard by everyone.

> *"when you see* [the robot] *at the reception, you don't want everyone to hear* [the conversation]"

Ease of use of the robot and ease of learning to use the robot (SUS)

Regarding the question on the ease of use of the robot, and if there are many things to learn to be able to use the robot, the participants found that the robot is easy to use, where you just have to talk to it, without pressing any buttons, and without the need to learn any particular functionality.

> "*It is easy [to use the robot] since you just have to talk*"

> "*I didn't press a button, yeah, I just have to adapt [to the robot]. Just like with a person in the end*"

Regarding the question about needing help to use the robot and the fact that most patients could learn to use the robot quickly, the participants noticed that the robot did not understand certain words, and instead, activated the answers with certain words. They think that a patient in the hospital would not be able to know which word to use to get the answer to their question. For them, a patient should not have to think about a specific word to use, he would use a vocabulary that he would use in his daily life, and be understandable by other humans.

> "*Your colleague, at one point, he just said the word 'QUIZZ'. There are certain words that need to be said, but the person who comes in doesn't know that those are the words. So that's the problem, you always have to enter into the logic of the information that you have put. And it's not the logic of the visitor, of the person who comes*"

> "*When the robot talks about a 'quiz', you have to take up this theme, but if you change the term to 'game', you have the impression that it* [the robot] *is lost*"

Suitability of the different services (functions) of the robot (SUS)

Regarding the question about the different services of the robot, most of the participants thought that the answers were too fast and too long. Indeed, some answers included too much detail, too much information, and the flow of words was too fast. This made it impossible for participants to retain this information and use it.

"*You turn left, etcetera. There, it should be slower, a little bit slower when there is path [to be given as instruction], especially when it is long*"

"*The length of the sentences. Yes, the amount of information that it gives you is the somewhat general misunderstanding that occurred. That's it, it's a little bit at all levels of the dialogue* "

Opinions on the inconsistencies in the robot's behavior or services (SUS)

Regarding the question about the inconsistencies of the robot, a large majority of the participants found the answers to the questions asked to be inadequate. Indeed, in some cases, the robot did not answer the questions asked and instead returned to the topic of food, which could cause confusion for some participants.

"*The robot's answers are not adapted*"

"*When it changes the subject of conversation and answers something other than what you ask is embarrassing.*"

"*I have the impression that the robot is limited in what it proposes*"

"*There is this problem of the topic of food that keeps coming back [regardless of what one asks the robot].*"

"*All the questions related to general care, it didn't answer. Either it didn't understand or it doesn't have an appropriate answer*"

Confidence and ease of interaction with the robot (SUS)

Regarding the issue of the difficulty of talking and behaving naturally with the robot, and the issue of trust in the presence of the robot, participants thought that the malfunctions of the robot may lead to a loss of patience. It could cause them not to use the robot in the long run. A loss of trust could also occur, as patients would no longer come to use the robot, knowing that it might not answer the question asked. During the test, the multiple malfunctions did not allow the participants to consider a trusting relationship with the robot, and especially felt that they could not test it.

"*When you come to a hospital, especially to consult, there is always stress, there is always stress despite yourself if you are healthy. And so, it has to be as easy as possible*"

"*I didn't really feel like I could test it*"

E. Appreciation of the dialogue capabilities of the robot (HWU questionnaire) - Wave 2

Participants were also to rate the robot's voice clarity, the quantity of information at one time, the robot's ability to "hear" what is said, Its understanding of what is said, the authenticity of the conversation and the robot's response time using a Likert-scale *(1= Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly agree)*. Overall, the results illustrate a diversity of positions with respect to the overall experience of interaction with the robot, with both positive and negative attitudes toward the robot. Globally, the participants were satisfied by the robot's voice clarity(m=4,7 ; ET=0,48 ; med=5) and the robot's response time (m=2,6 ; ET=1,04 ; med=3) but they found that too much information was provided at one time(m=3,8 ; ET=1,64 ; med=5) with great disparity in this item's answers. The results were mitigated regarding the robot's "hearing" ability (m=3,1 ; ET=1,38 ; med=3) and the natural aspect of the conversation (m=3,2 ; ET=1,34 ; med=3) but they seemed to agree on the fact that, overall, the robot failed to understand the meaning of the participants requests (m=2,6 ; ET=1,04 ; med=3) (see Figure 15 for a graphical illustration of these results).

Figure 15 displays the mean scores given by the 13 participants in the different dimensions assessed regarding the robot's features of the robot (HWU questionnaire).

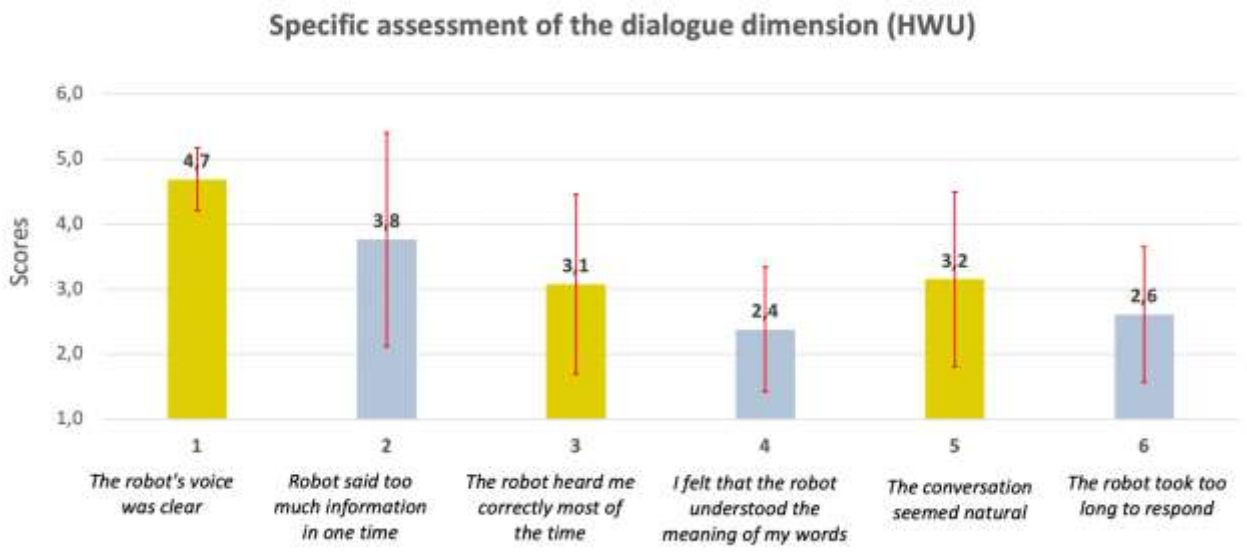


Figure 15. Mean scores on the different dimensions assessed regarding the robot's dialogue capabilities - Wave 2 (n=13 participants)

Below we present some of the qualitative feedback given by the participants regarding the robot's dialog capabilities in this experience.

The clarity of the robot's voice

Globally, the robot's voice was considered to be clear.

> "*I am hard of hearing so I understood it well. I found it clear, consistent*."

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

"There, I understood. I understood everything so"

The subtitles on the screen allowed reading at the same time, although there were some errors in transcribing some of the participants' words.

"It shows up on the screen [the dialogue] and that's good"

Feeling of being heard and understood by the robot

Overall, the participants did not feel that they were understood by the robot in the dialogue. The robot did not recognize some words used by the participants or did not succeed in recognizing the required request, its answers were not always adapted, or too long. Sometimes the robot answered too early and thus cut off the participants' speech.

"This time I would say no [the robot didn't understand] because sometimes it's really messed up"

"I would be pretty hesitant [about feeling understood by the robot] huh"

"*I don't know if it heard me and I don't know if it understood me"*

"*I didn't finish my sentence, it was already answering"*

"When I spoke briefly, the robot responded well, but when I spoke at length it did not"

Perception of the naturality of the dialogue with the robot

Opinions on the matter were rather mitigated, several comments were made by participants describing not having felt the verbal interaction with the robot in a natural way.

"The interaction [it is not natural]. No, no, it [the robot] is dictating"

"So not at all [the robot was not natural] Me, I was natural and it, the poor thing, it comes out what they put on it"

Some participants were uncomfortable with the familiar tone of language used by the robot.

"One thing that shocked me was when the robot said 'Hi' [instead of Good morning] "

"There's everything that's good but it [the answer] should be phrased politely."

About the robot giving too much information

The information provided by the robot was considered too long and detailed in some cases, especially with regard to orientation instructions to find a place. The participants said that too much information was difficult to deal with and to remember.

> *"When you ask for the way, it will indeed give you the way, but you should ask yourself the question, will the human person record it all?"*

> *"A few times, yes, the robot said too much information at once."*

The robot's response time

For some participants the response time of the robot was acceptable, for others, the robot took a long time to give its answers to the questions asked by the user.

> *"Not all the time. But it's true that there are times, yes [the robot takes a long time to respond]."*

> *"Sometimes it's very responsive, and then sometimes...it's looking for [the answer]."*

> *"For the most part, it responded quite quickly. But let's say, 2 times, it must have hesitated"*

F. Semistructured interview on the ethical dimension - Wave 2

This section presents the main opinions given by the participants regarding the ethical issues addressed in the semi-structured interview (see annex 4). Some participants' verbatims were extracted from the interview transcripts and used to illustrate different views.

1/ Opinions regarding the development and use of robots with human-like appearance (humanoids)

Several participants said they appreciated the fact that a robot that is going to be used as a receptionist has a human appearance. Since the main role of this type of "receptionist robot" is to give information to the user, participants thought that human attributes could contribute to a better acceptance of the robot.

> *"If the robot it's a receptionist, it's still more user-friendly [that the robot has a humanoid appearance]",*

> *"It's more fun and more attractive [that the robot has a humanoid appearance]"*

> *"It's more encouraging to address a robot that looks like a human being"*

> *"I think it should be a robot that is as human-like as possible. It's more encouraging to address a robot that looks like a human being"*

However, in some cases, the acceptance of the humanoid aspect of the robot, related to the role of receptionist demonstrated during the evaluation, was accompanied by a certain frustration and resignation on the part of some participants.

> *"Why not,[that the robot has a humanoid appearance], since basically, we're asking a robot to replace a human",*
>
> *"I think it's even better [that the robot has a humanoid appearance] than just having a cube with a voice that talks. It gives at least a playful touch"*
>
> *"We have to deal with it (the robots). We don't have a solution, even if I say no (I don't want them) anyway there will be. So the best thing is to be prepared."*

Nevertheless, other participants declared that the robot should not be totally human-like since too much realism in the design may lead to confusion and entail a certain risk of deceiving the potential users of the robot.

> *"When the robot is very close to a human, it's confusing,"*
>
> *"Ari doesn't [...] confuse me because its [design is] neutral enough not to be confusing. You can see that he's an automaton, so it doesn't bother you"*
>
> *"It [the robot] shouldn't be too realistic. It should not be an identical replica of the human, especially if it is to be used in public places"*

Regarding the aspect of the robots, other participants declared that the tone of the robot's voice was as important as its physical appearance: *"We'll say the audio form is just as important".*

2/ Opinions regarding the social robots in the hospital, advantages and drawbacks

Concerning the use of social robots in hospitals, the results show that most participants considered that robots in the hospital should be supplementary to care workers and that they should never replace humans:

> *"It [the robot] can be a good thing [...], but it should not diminish other services, it should be an addition that can help".*
>
> *"It must not be to replace the human being. A robot that comes to help a caregiver lift a person who has fallen, I agree with that, but a robot that comes to do the consultation..., no way"*

A few participants emphasized that the use of the robot should focus on repetitive and painful tasks for humans.

> *"The robot is not asked to think, it is asked to execute,"*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

"The robot can be a help, really, but for very simple things, giving directions, what documents to provide [or] how to pay".

"I think it's a plus, compared to the problems of recruiting staff, even for carrying meal trays, it doesn't shock me"

Other participants expressed some hesitations regarding the use and the usefulness of robots in hospital environments.

"It is attractive, [...] but at the same time, it certainly does not replace a receptionist."

"For sick people who arrive at the hospital being greeted by a robot is not necessarily reassuring, but that in the context of regular consultations the robot would not generate any anxiety."

Overall, the interviewees found the robot interesting and the integration of these new tools in the hospital is synonymous with progress and modernization.

"A robot like ARI is excellent"

"[the use of robots] is a normal evolution"

"It [the robot] can be a precious aid, of help and assistance, but that it should be used with moderation."

Finally, the question of the target audience to whom the robot would be best suited, in the context of a hospital, was also discussed.

"In the children's department, it would be very good. For the older people I don't think so much"

3/ Opinions on the data collection and treatment made by robots

Regarding the question: "Do you think it is right that robots record and store information?" ; the vast majority of participants thought that everything depends on the purposes of the robot's use of the data and the way in which the confidentiality of the users' data is respected:

"It depends on how the data is used. If it's done to improve the service, there's no problem. If it's to put my data on the internet, no"

"[I'd agree with the data collecting and storing data only] f this data on the robot is protected"

"It depends on how you use the data"

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

"if "it's collected online, no. [...] If it's protected, like with computers, it's fine"

Many compared the data collected by the robots, in the context of the hospital, to the data recorded in medical registers (paper or digital files) or related to social security. The fact that the proposed use of the robot takes place in a hospital, an environment widely accepted as serious and reliable, makes it easier for many participants to accept data collection by a robot in this context.

"I don't have a problem with that [the collection of data by the robot] as long as they don't ask me for my bank account. I don't care about my health data, it doesn't bother me. Social security knows everything about you"

"Every good doctor consults his files, now he consults his computer, so I agree [with data being collected by a robot in the hospital context]"

"Data is anyway collected whether it is recorded by a robot or by the hospital"

However, despite the acceptance of data collection by the robot, participants felt that the system should not save personal data in the long term:

"Beyond the conversations, it [the robot] should not keep the user's data"

"I think it depends on when you say you'll keep it, and how long. As long as we are informed, yes, I think it is even important,"

"Data storage would be a bit annoying."

4/ Concern about the mistakes (bugs, accidents) that social robots might make/cause

Overall, the participants did not express any particular concern about the errors that the robot could make, on the one hand, they were rather confident in the robot programming (reliability of the computer systems), on the other hand, the participants considered that in the case of using the robot mainly to provide information, the potential errors that the robot could make would not have important consequences, apart from being a little annoying.

"I didn't even think about it [the mistakes the robot might make]."

"A robot never makes a mistake. It only makes a mistake if the programmer made a mistake"

"A machine will always execute only the order it was programmed to do. What can happen sometimes is electronic failures"

"it can happen indeed [that the robot makes mistakes], but like any software, or any computer system, we must not worry about everything, we must progress.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

> "*it depends on what level they [the robots] intervene. If it's to give the information regarding the floor number, it's not a big deal*"

> *"If the robot doesn't orient well, it causes a delay, it causes irritation. But it's not like a medical intervention where it would be really dramatic"*

5/ Views regarding the possibility of having an emotional attachment to a robot

Regarding potential attachment to a robot, the opinions of the participants were quite diverse. For some people, such an attachment is not possible and they do not consider that there is any risk of developing an effective bond with a robot such as the one evaluated during the experience.

> *"The emotional bond can only develop from one living person to another living person. For there to be affection, there has to be a relationship, you have to be able to touch, you have to be able to feel"*

> *"I don't think so, because for me it [the robot] is a machine"*

Some people, while declaring that they do not feel concerned by an immediate attachment to the robot, do not exclude that this attachment could develop in the future.

> *"So that's the big question. No today, no. I don't think so. It may come one day but it's not going to happen soon"*

> *"I don't know [if I could develop an emotional attachment to a robot], eventually, I don't know…"*

> *"At the beginning, I was very enthusiastic about this [technological] development, because I told myself that it would free people from ungrateful tasks, however there is the risk that it goes too far. Too far, that is to say that it [the robot] could start showing some kind of empathy"*

For other participants, it could be possible to have a basic emotional bond with a robot, which does not seem to be of concern and can be observed in other situations.

> *"So, when we say emotional attachment, I think there is a range of degrees and so yes, I think there can be a part of it, like people who have an attachment to their car"*

> *"Like children [who] can find comfort in a robot"* and who *"will tend to see it as a plaything"*

On the contrary, other participants seemed worried about the attachment they might feel to a social robot. To mitigate this risk, emphasis is placed on the importance of the robot's appearance and behavior being easily identifiable as that of a robot.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

"I say yes [the possibility of attaching to the robot exists] unfortunately. In my opinion, you shouldn't have any attachment. We don't have an attachment to the robot that cleans the floor"

"It's even what worries me, while at the same time, it fascinates and pleases me because you get into an imaginary world as if you were playing with a doll, you know. That's why I'm more comfortable with a robot that remains a robot".

6/ Opinions on the risk of the replacement of the human being by robots

Positions on the risk of human replacement by robots varied, but most participants acknowledged that such a risk exists and discussed some of its consequences. Some recognized this risk but insisted on the importance of defining the role of the robot as complementary to human assistance.

"Yes, there is a risk [of replacement], I think it can happen in situations where we want to plan everything and do everything perfectly, and then someone says 'Oh, finally, for this job or this part of the reception we'll put a robot, and finally there are more robots than humans, yes I think it's a risk"

"On some forms, yes [I worry about the risk of replacement], whether the robot is a complement yes, it is , a replacement no it isn't"

"Yes [the risk of replacement exists]. The robot must remain in its place"

"From a social point of view, yes, it is sometimes a bit worrying, [the risk of replacement] but robots are useful in industry, to accomplish repetitive tasks and why not in a hospital? Yet the staff at the reception desk will be unemployed if they are replaced by robots.'

Nevertheless, there were a few participants who stated that they did not believe that this replacement of humans by robotic tools was possible at all.

"Robots cannot replace people, they can only help. It's not the same thing. A machine will never replace a human, it's impossible"

"It is not possible to replace humans. In any case, humans are needed to program them, so robots can't replace them"

Some participants acknowledged this risk of replacement, and accepted it in a rather positive way, seeing it as a solution, since the current situation of labor crisis in health care facilities is not going to improve. In a more pragmatic way, these people expressed a degree of acceptance of the replacement of the human by the robot.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

"I see it as a positive thing [the introduction of robots]. We have to find solutions so that people can have access to medical spaces. In a few years, we won't be able to find people to help the elderly"

"Yes [the risk of replacement exists]. It's more profitable [to have robots than humans], they don't make demonstrations, they work 24 hours a day"

3.2.3 Study 1: Wave 3 experiment (Living Lab - Simulated environment)

Period

March 2023. The experiments took place over 2 and a half days, at the Broca Living Lab.

Objectives

The main objective of the Wave 3 experiment was to evaluate an updated version of the architecture working on the ARI robot (version October 2022) with potential users of the solution (older adults using hospital services and their accompanying persons) in a simulated environment (i.e. in living lab conditions). Wave 3 assessments were structured to:

- Continue the process of user data collection and its provision to the technical partners in charge of the development and improvement of the technical modules of the robot (vision, navigation, dialogue, emotional and engagement aspects).

- Conduct further collection of user feedback regarding acceptability, usability and satisfaction with the SPRING robotic solution.

- To evaluate the multi-party dialogue capabilities of the SPRING system, by proposing collective interactions (two users at the same time) with the robot.

- Conduct further exploration of the opinions and perspectives of potential users regarding the ethical dilemmas involved in the use of social and assistive robots in a health care delivery setting.

Participants

A total of 20 volunteers took part in this assessment. Volunteers were recruited through senior associations and former users of the hospital. After having obtained the agreement of the older adults to take part in the study, the researchers arranged a meeting with the participants for the test session in the living lab. The socio-demographic data corresponding to the participants are presented in Table 8.

Table 8. Socio-demographic data for participants in the "Wave 3" Living Lab assessment

Variables	Modalities	Total
Count	-	20 (100%)
Gender	Male	4 (20%)
	Female	16 (80%)
Age, years	*Min : 67 ; Max : 89* *Mean:* 78; *SD: 6,15*	

	[67 – 70[[71-75[[76 – 80[[80 +[	04 (20%) 05 (25%) 07 (35%) 04 (20%)
Living Arrangements	Alone With Partner	14 (70%) 06 (30%)
Education Level	High School Diploma College Degree	08 (40%) 12 (60%)

Material

The robot ARI (PAL robotics) configured using the different modules developed for the SPRING project was used as a platform for the assessments. A global description of the ARI robot and the specific modules developed for this platform in the SPRING project have been described in Section 3.2.1 (Material) of this report.

The specific configuration of the robot for the Wave 3 assessment comprised the following features:

- Dialogues were handled automatically by the robot in real time. (no Wizard of Oz)
- Individual and multi-party interactions were proposed, meaning that participants performed individual or collective (two persons) verbal interactions with the robot. The robot was provided with a female Text-to-Speech (TTS) voice (Acapela's UK English voice 'Rachel') set at 50% of the maximum volume). A series of tasks, related to reception and orientation for hospital users, was developed for this assessment.
- Dialogue subtitles were displayed on the robot's screen,
- The robot moved towards the user at the beginning of the interaction and maintained its orientation towards the user during the interaction.
- Not arm movements of the robot were introduced
- The robot's gaze was always directed towards the person

Procedure

Participants were invited to enter a room where the Ari robot was located. The experimenter explained to the participants the different implemented use-cases. In this set up the robot was programmed to come to the participants and greet them. The robot was programmed to greet and introduce itself to the participants automatically. The participants were asked to play the role of a patient coming to the hospital for the first time and asking for information regarding their visit. They were requested to ask questions to the robot one at a time. After this interaction phase, the researcher proposed a multi-party themed interaction where the participants were asked to

enunciate questions simultaneously, one playing the role of a patient and the other one playing the role of someone accompanying the patient.

The first stage of the interactions was constituted of 'open' dialogues (i.e. participants were free to choose their own enquiries). The other one was fully task-based, which meant the participants were supplied with a set of consecutive tasks that had been designed to illustrate the key functionality of the system. The individual tasks were represented as images (see Appendix A for an example task sheet) to avoid prompting participants to use particular phrases when talking to the robot.

Table 9. Examples of multi-party dialogue contents managed by the robot

Type of question asked by the user	Robot responses
User 1: What paper do I need to register at reception? User 2: Do I need my health card to register / check in?	To user 1: You need your convocation and your vital card. The secretaries will explain everything to you. They will help you if you forget. To user 2: Yes, your health card is necessary for your registration on your arrival. Did you take a ticket?
User 1: Where can I wait ? User 2: Can I have a paper and pen?	To user 1: For you, You can wait in the waiting room, the nurses will call you. To user 2: And for you, A nurse can help you, ask them!
User 1: Do I need to pay something?	To user 1: No, your consultations are covered by social security. The secretaries will be able to inform you better than me.
User 2: Can I get / I need a wheelchair / cane / walker to get around?	To user 2: A nurse can help you, ask them!

Figure 16. Example of interaction between the participant and the robot in the experiment Wave 3

Assessments

At the end of the interaction task a comprehensive set of assessments was conducted to obtain a complete overview of the users' experience and their perspectives on the robot and its usefulness. These assessments included:

(A) Acceptability of the system was assessed using the acceptability questionnaire (AES) was administered. For each of the 6 items on the scale, the participant was invited to argue his response (scale in Annex 2).

(B) Participants were then asked to answer two questions regarding their subjective experience of the robot:

- Rating the *appearance/design* of the robot (using a 5-point Likert Scale and qualitative feedback)
- Rating his/her *feeling about the interaction* with the robot (using a 5-point Likert Scale and qualitative feedback)

(C) Participants were invited to provide some examples of improvements that could be made to the robot, if relevant to the participant (using qualitative feedback)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

(D) Usability of the system was assessed using the System Usability Scale and qualitative feedback was requested from participants to explain the notes attributed in the different dimensions of the scale (see Annex 1).

(E) Finally, we conducted the semi-structured ethical interview with the participants (see Annex 4).

Results

(A) Acceptability of the robot - Wave 3

This section presents the results obtained in the evaluation using the AES (Acceptability Scale).

Overall, the results illustrate a diversity of positions with respect to the overall experience of interaction with the robot, with both positive and negative attitudes toward the robot as expressed by the wide range of the observed standard deviations for each dimension suggesting that their interactions were either acceptable or not acceptable. The highest scoring dimensions were the ease of use (m=3,9 ; SD=0,91 ; med=4) and the robot comprehensibility (m=3,75 ; SD=1,07 ; med=4) followed by the perceived usefulness (m=3,5 ; SD=1,28 ; med=3) and the response time (m=3,35 ; SD=1,18 ; med=3). As in waves 1 and 2, the lowest scores are reported on the overall satisfaction (m=2,45 ; SD=1,15 ; med=3) and the enjoyment regarding the use of the robot (m=2,45 ; SD=1,19 ; med=2).

Figure 17 presents the mean scores for each dimension of the acceptability scale for the whole sample (n=13 participants) who took part in the Wave 2 assessment.

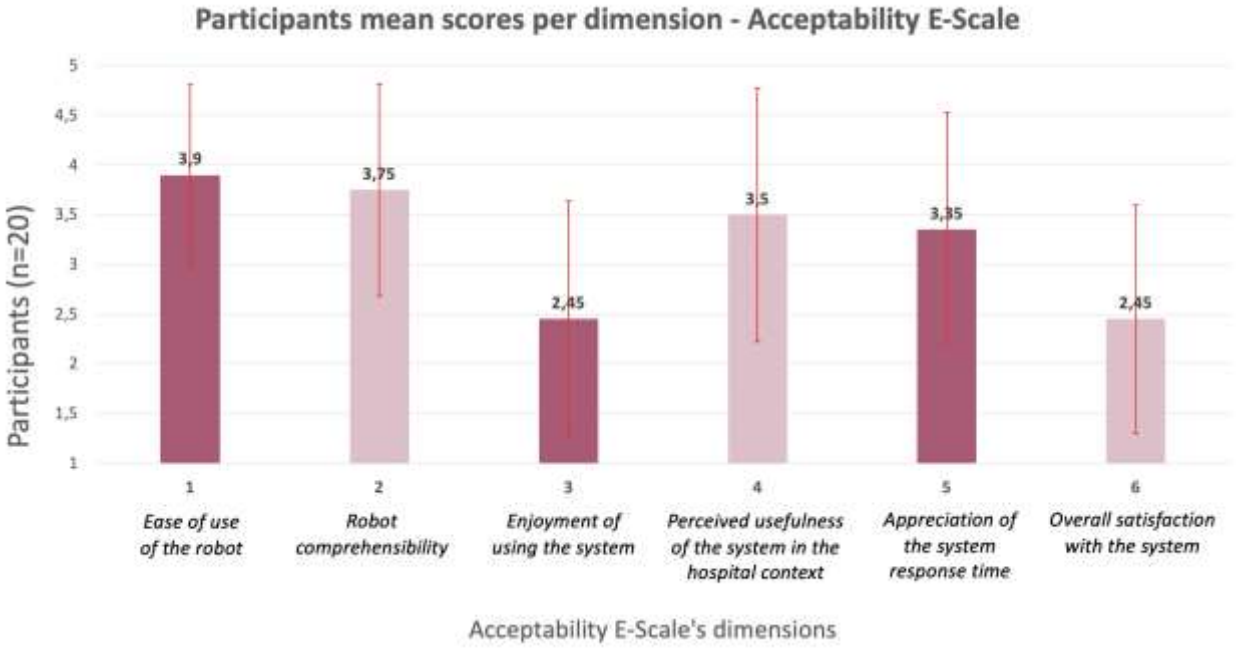


Figure 17. Mean acceptability scores (AES) for the assessments in Wave 3 (n=20 participants)

Participants in Wave 3 reported a mean total score of 21,7 (SD=0,81) over 30. This sample median was 20 over 30. Acceptability E-Scale cut-off is set to 25.8 over 30, indicating a satisfactory acceptability level, and no participant reported scores above the threshold. Wave 3 participants' scores ranged from 19 to 28 over 30. 45% of the participants reported a score lower or equal to 20 and 55% reported a score between 20 and 25.8

Figure 18 presents the mean scores on each dimension of the AES for the whole sample in Wave 1 (n=20 participants).

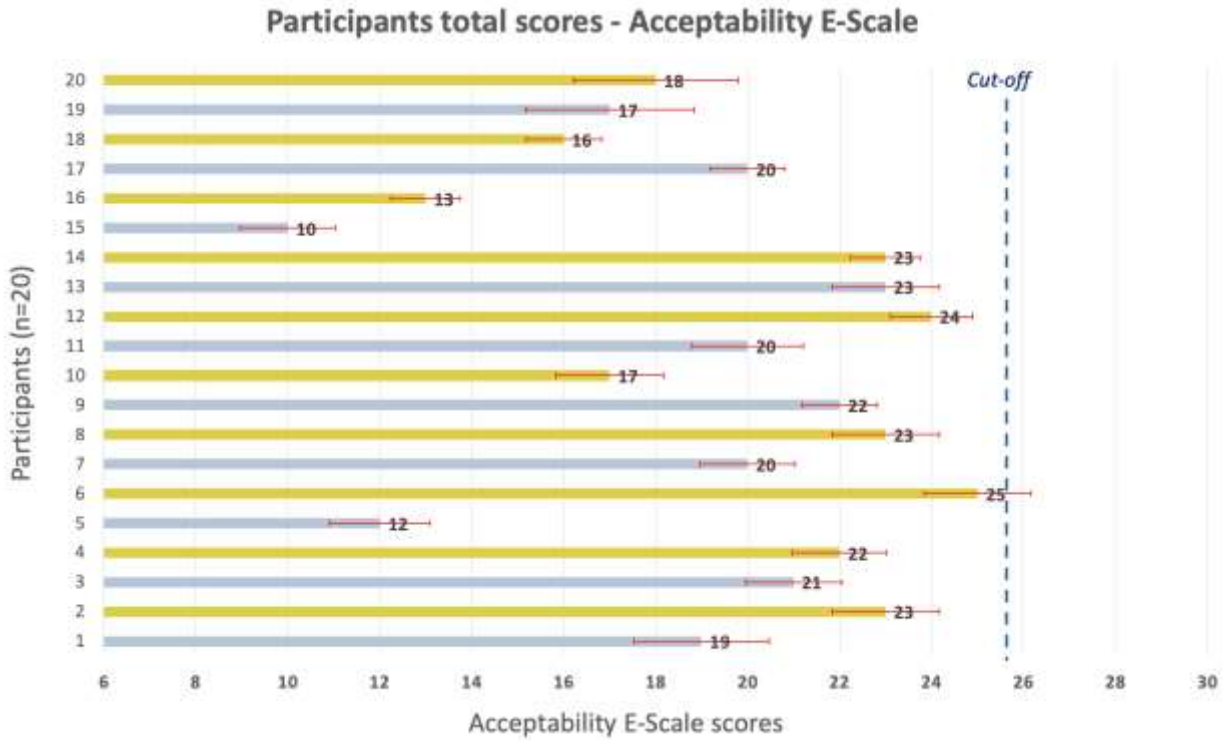


Figure 18. Global scores (AES) for the assessments in Wave 3 (n=20 participants)

Below is presented an analysis of the qualitative returns given by the participants during the evaluation of the dimension of the acceptability of the robot in Wave 3.

Ease of use of the robot (AES)

For most of the participants the use of the robot did not represent a particular effort.

> *"You don't need any help to use the robot".*
>
> *"you just have to talk to it."*

For other people, technical problems encountered when interacting with the robot result in it not being very easy to use.

> *"It doesn't work well, I give it a neutral rating [neither easy nor difficult to use]"*
>
> *"[neither easy nor difficult to use] it has trouble, it stammers".*
>
> *"[neither easy nor difficult to use] it has a recognition problem".*
>
> *"It did not respond to what I wanted. I had to rephrase"*

Robot comprehensibility (AES)

Overall the participants found the robot comprehensible in terms of hearing and diction but the content was sometimes incongruous. In this sense, the robot did not provide a sense of dialogue with the users. The users perceived non-adapted answers, not taking into account the previous exchanges or the information the users had already given.

> *"It was understandable"*
>
> *"I was surprised because I was expecting the answer and it wasn't the right one. But it was easy enough to understand".*
>
> *"The robot was understandable but did not answer the questions appropriately".*
>
> *"The [robot's] answers did not match the questions".*

The addition of the written transcript of the speech, on the robot screen, was appreciated.

> *"As it was speaking, it was scrolling and you could read it on the screen, that was good. From the perspective of a person with limited abilities, sound and sight at the same time, it increases concentration".*
>
> *"Problems with voice intonation. You can read, so I understood, it completes. It's pretty easy to understand because of the screen".*

Some of the answers provided by the robot were too long, and some topics seemed very repetitive in the robot's discourse

> *"The robot's speech was too long or complex, and difficult to understand orally without visual support".*
>
> *"Some of the answers, especially the one about the cafeteria, were too long. The theme of food came up often, regardless of the question asked"*

Finally, for some participants, the multi-party dialogue component in the test scenario was rather confusing.

> *"I have to admit that it was pretty confusing. I think there were two of us [interacting with the robot], so it would have been better if it was one person by one person. It was difficult".*

Enjoyment in using the system (AES)

For some of the participants the experience of interacting with the robot was described as positive. Among the positive aspects mentioned by the participants that contributed to making the experience a pleasant one include:

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

“It was a bit like a game [the experience with the robot]”

For quite a few participants the interaction experience was not entirely pleasant, for several reasons. The absence of real interaction was cited as one of the reasons why the experience with the robot was not fully described as positive.

"I did not find the experience enjoyable so far”

"I didn't feel like I was interacting with the robot"

“I'm going to be very severe, it gave me two good answers, and for the rest it went wrong”

Many participants were confused by the robot's answers, which did not regularly correspond to the question asked. Many felt that the robot did not understand them. This type of problem greatly affected the quality of the experience and the participants' perception of the robot.

"I made sure to ask simple questions, where are the elevators. The robot didn't know. It's like we don't speak the same language."

"At first when we started [the interaction] the robot suggested to go to the restaurant when we were hungry, when that wasn't our question."

"It doesn't answer enough questions."

“We had to adapt to it. It was the first time”

Usefulness of the robot in the hospital context (AES)

Overall, the participants found that the robot could be of moderate utility in the hospital setting, for example for the provision of various services, for instance, to help with the reception at the hospital: *"It could be useful for the welcoming, to show the way". “it could be useful when it is well programmed”.* The usefulness of the robot in this type of environment will depend in any case on its reliability:

“it could be useful when it will be properly programmed”.

"It [the robot] could have its use but for the moment it is very limited. That is to say, for general information, it can be useful, insofar as at the reception desk when you arrive, there is no one there. So in the meantime, it can be useful.

The lack of staff appears to be a deciding factor for the use of the robot in hospitals. o assist healthcare staff *“If there is a large number of people in the hospital, it can help out. It can help the staff”.*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Some participants consider that today the integration of robots in the hospital will not be well accepted by the users, especially the older people.

> *"I think it [the robot] would be fun for the kids but today I think the majority of the population is not used to this kind of device at all. So in the immediate future, I wouldn't be sure it would be useful."*

Other participants fear that robots will replace humans:

> *"I am reserved because I am afraid that it will replace humans. That's my big concern. The human relationship..."*

Appreciation of the robot's response time (AES)

Some participants considered that the response time was acceptable: *"This slowness characterizes a relationship with a machine."* Other participants thought that the waiting time for an answer was also long:

> *“When you are in front of the robot, you are so used to a robot being fast, knowing everything, it is quite disappointing. It's not fluid”*

> *“it should be explained that it has difficulty in recording. Sometimes the written sentences are not the same as the ones we say. And it takes a long time to answer.*

> *“You have to let the robot answer, and we are used to going fast and getting answers quickly, so we don't expect the robot to think about giving an answer”*

For some people, even if the response time seemed to be correct, the problem of inconsistency in the robot's answers was mentioned: *“Yes [the response time was acceptable], but it doesn't respond to our requests”.*

Finally, some participants mentioned that they had to repeat their question and were then cut off by the robot, resulting in a negative user experience.

Overall satisfaction with the robot (AES)

The overall experience of the participants in this evaluation was rather mitigated, some positive points were mentioned as contributing to a positive experience. However problems with the consistency of the robot's dialog significantly affect the quality of the experience and, consequently, user satisfaction.

> *“I am not satisfied [with the robot]”*

"If they build machines, it's to make them perform better, even better than we do as humans"

"There are positive and negative points, I am neutral"

"We have to make the difference between the game, a device that we take for a game. And then, precisely something that will serve us. You have to make a difference. For me frankly, I am neither satisfied nor dissatisfied [with the robot]"

"Physically, it [the robot] is nice. After that, it has to work. A person who is upset, can become very aggressive because it does not respond"

"What bothered me was that it talked too much about food, it was all about food. Whatever the question, it comes back to food"

B. Subjective experience of the robot

Participants were asked to rate the appearance (aspect, design) of the robot using a Likert-scale *(1= Very unpleasant, 2=Unpleasant, 3=Neither pleasant nor unpleasant, 4=Pleasant, 5=Very pleasant)*. Mean score for the 20 participants was 3,9 (SD = 0,91, median=4) (See Figure 19). Overall, the robot's appearance is rather appreciated. However, we also find some opposite views with respect to the actual design of the robot.

Participants were also asked to rate from 1 to 5 their feelings during the interaction with the robot using a Likert-scale *(1= Very negative, 2=Negative, 3=Neither negative nor positive, 4= Positive, 5=Very positive)*. Mean score for the 20 participants was 3,7 (SD = 1,03, median=4) (See Figure 19). Overall, the results illustrate a diversity of positions with respect to the overall experience of interaction with the robot, with both positive and negative attitudes toward the robot. We did not find a representative overall trend in this regard.

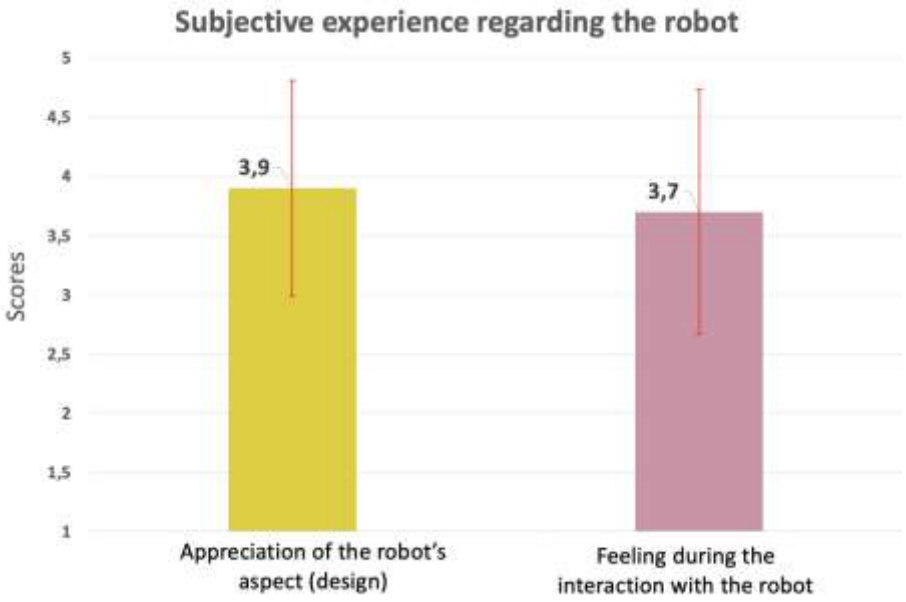


Figure 19. Mean scores of the assessment regarding the subjective experience during the interaction with the robot - Wave 3 (n=20 participants)

Opinions about the appearance of the robot were varied, with a few positive comments mentioned:

"The face is nice, it's gentle. It is feminine this robot, for me it is feminine"

"It was nice, yes. I like his eye movement."

The fact that ARI was clearly identified as a robot was considered a favorable point:

"You don't want it to look like a human, that would get serious for me. It's okay, it's friendly".

"It's a robot is all"

"You can tell it's a machine, so that's good."

"Embodiment is still important, but not too important"

Other participants were less satisfied with the design of the robot citing different reasons:

"It is too impersonal". "It is too rigid".

"It's a bit of a long dress. Maybe it should be a little more manly?...I don't know".

None of the participants noticed the change of color of the robot, when it listened, or when it spoke.

C. Usability of the robot - Wave 3

With respect to the evaluation of the usability of the device, the results of the evaluation made with the SUS scale showed that

In wave 3, the participants reported a mean score of 66 on the SUS scale (ET=17 ; med=65 ; med=71) which would be considered below average but still "OK" according to the SUS rating scale (refer to Figure 4, P5). As a reminder, a SUS score above 68 would be considered above average and anything below 68 is below average. Participants 15, 16 and 18 can be considered as outliers and not taking their scores into account would significantly change the mean score (m=72 ; ET=9,9 ; med=72,5) making the mean score considered as "good" on the SUS rating scale..

Figure 20 presents the scores given by the 20 participants on the SUS usability scale.

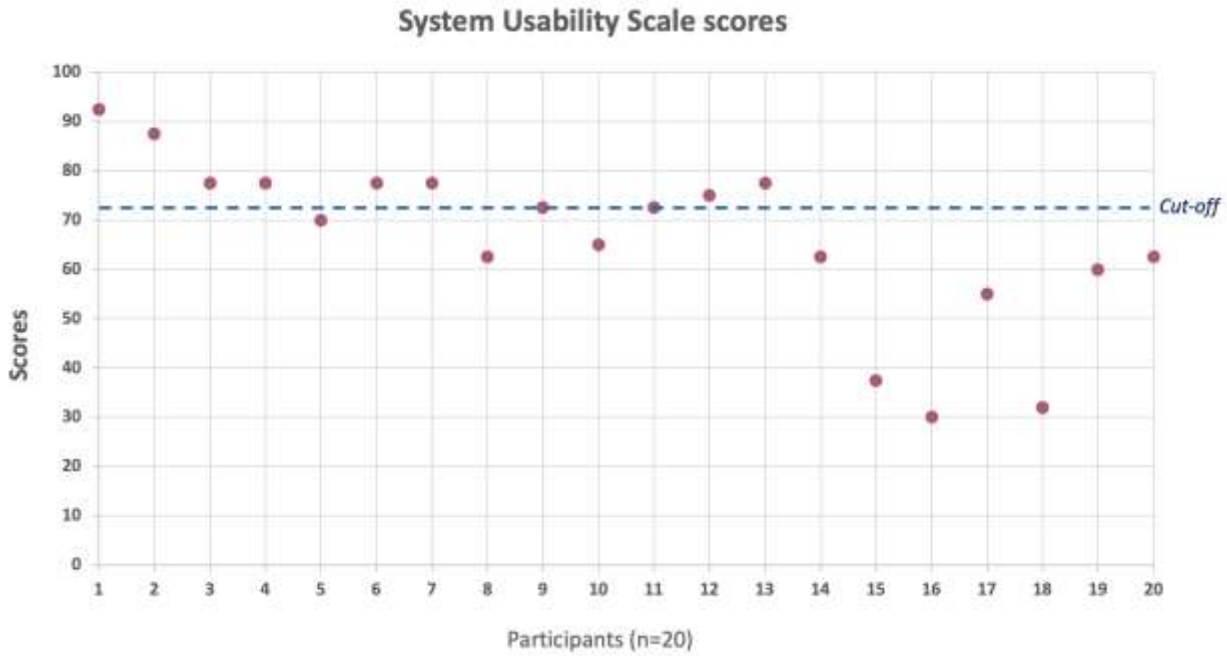


Figure 20. Scores for the scale of usability (SUS) for the assessments in Wave 3 (n=20 participants)

Interest in using the robot regularly in the hospital (SUS)

Participants agreed on the potential usefulness of the robot. However, they emphasized that they would prefer to interact with humans rather than with a robot, if the choice is given.

"If the secretary is already busy with three people, the robot can be useful".

"Yes, I think it can be useful, if there is a rush it can double the reception".

"It can be interesting if it exists and works well. If there is a human, I will go to the human. If there's no one there, I'll go to the robot."

"No, I'm not very interested. The first time will be fun, but I'd rather talk to a person."

Perceived complexity of dialogues with the robot (SUS)

Concerning the question about the conversations with the robot being unnecessarily complex, some participants said that the robot did not answer the questions properly. They also thought that the answers were too long.

"It's a weird sentence [what the robot says]. It's not complex, it's inadequate".

"It depends on the answers and the questions asked. Sometimes it's well answered, and sometimes it's not".

"You have to repeat [to the robot], it lacks fluency. A person who is anxious, who doesn't have an answer, it can make them angry".

Ease of use of the robot and ease of learning to use the robot (SUS)

Regarding the question on the ease of use of the robot, and if there are not many things to learn to be able to use the robot well, the participants found that the robot was easy to use since you just have to talk to it.

"It's when you don't get the response you expect, that's when you don't know what to do. So it's a little difficult"

Most participants said that they did not need help to use the robot and could learn to use the robot quickly. However, some people noted that a little help was needed at the beginning of use to become familiar with the robot

"For now, yes, help is helpful... But you get used to it. I think eventually we won't need help. Eventually you will get familiar with it"

However, they thought that persons with cognitive impairment would have difficulty using robots.

"It's easy to use except for the people in the geriatric hospital. If they are not accompanied, not everyone could use it. There are people who would have significant difficulties, physical or psychological."

"If I'm fine, I can interact with it [the robot]. Now if I have significant cognitive problems, it will be something else.

"If it's someone who is hard of hearing or has significant problems, I don't know if they're going to engage with the robot".

Suitability of the different services (functions) of the robot (SUS)

Regarding the question about the different services of the robot, most of the participants thought that the answers were too long. Indeed, some answers included too much detail, too much information.

"It's not complex, but it's too long. The answers are too long. Something more synthetic would be better".

" It answers the question but it puts a huge amount of information on us, for the cafeteria. It explains everything there is, but we know what there is in a cafeteria. Drinks, food. The answer is too long".

Participants would like the robot to list its available functionality.

"It [the robot] is unable to list its functions".

Several participants did not like the robot's voice:

"The voice is too hollow".

*"It's very stereotypical. The tone of voice is very robotic. There's no emotion. It's choppy. It's a robot language, very flat***"**

Opinions on the inconsistencies in the robot's behavior or services (SUS)

Regarding the question about the inconsistencies of the robot, a large majority of the participants found the answers to the questions asked to be inadequate. Indeed, in some cases, the robot did not answer the questions asked, and instead talked about food.

"There is no continuous exchange, like with a conversation partner. You ask a question and you get an answer. There is no continuity"

"You ask it [the robot] different questions on different topics but it only talks about food"

"You ask it [the robot] where the doctor is and it offers you a glass of water"

Some participants seemed reassured by the fact that the robot has inconsistencies.

"A robot can only be incoherent, otherwise it would be like Blade Runner".

Confidence and ease of interaction with the robot (SUS)

Some participants thought that interacting with the robot could be difficult. They did not behave naturally with the robot because they tried to modify their ways of behaving to have a satisfactory interaction with it without success.

> *"It's not easy to be natural with him. I thought you have to say the words in a certain way, and even speak up"*

As a result, people did not trust the robot's responses.

> *"I don't trust its answers"*

On the other hand, some participants thought that interacting with the robot was easy because the robot always responded calmly.

> *"I wanted to say that there is something nice about it. It's that with humans you can have someone in a bad mood, and that's normal. Here you are with someone who is very calm, who can calm us down. He can repeat things over and over again without getting upset"*

Some participants personified the robot and attributed positive intentions to it.

> *"Yeah, it [the robot] wanted to help us so much, it didn't help us".*

D. Suggestions for the improvement of the robot or the robot-related services - Wave 3

The participants provided some suggestions to improve the interaction with the robot.

- The robot's voice was considered "too robotic", and that "lacked emotion". *"The hollow voice is unpleasant to me so it's something to be modified"*
- Modification of certain aspects of the robot head: *"The human appearance is nice. But it would need eyebrows"; "I wish it had hair" ; "If we embellish it a little bit [the design of the robot], a little bit lighter, it can be nice".*
- Better adapt the robot's gaze to the user: *" It has to look at the other person, otherwise I don't know that it's talking to me"*
- An improvement of the robot's reaction time is desired: *"If we already improve the waiting time [to get the answers from the robot] I think it will be okay, we will get used to it".*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

E. Semi-structured interview on the ethical dimension - Wave 3

1/ Opinions regarding the development and use of robots with human-like appearance (humanoids)

Some participants said that they preferred the robot to have a human form as it provided more pleasure, more fun and more reassurance than with a robot with another form.

> *"Yes, it's nice, it makes you want to talk to him more than to a robot in the form of a cube".*

> *''Yes, it's even reassuring. Otherwise, we wouldn't even go to him".*

> *'' For me, yes. It's mandatory after all. I can talk to my phone for fifteen minutes, but to deal with material things, it's easier when it looks human because it makes you smile. That's why I want to put hair on ARI".*

Some participants accepted a humanoid form but wanted the robot to be differentiated from a human:

> *'' I don't approve of the human form. ARI has a human form, but you can see that it's not a human, I prefer it that way".*

> *''Yes, I prefer to see that it is a robot. It can have a shape that looks like a human, as long as you see that it's a robot. It's nice, we go to it. We know it's still a robot".*

> *'' I would be opposed to it looking like a human. But ARI does not look like a human. To look like a human would be to have hair, and a mouth, not eyes that blink or ears that light up like ARI. A robot that looks like a human to me is a problem, it leads into another world. It's worrying, I'm not from the science fiction generation. Maybe it feels different for other generations".*

> *''For me there must be a difference between the human and the machine: you can identify that it is a machine. But if it's completely human, I don't like it. But I like what ARI looks like".*

For other participants, the fact that the robot had a human or humanoid form was not relevant:

> *"I don't see the point of the resemblance to a human, we know it's a machine, from then on I don't mind. It could look like a pet, it would be just as nice".*

> *"At the entrance to the hospital, you could have a kiosk, or even an animal that would answer. It's not unpleasant, and if it's done well, it's pleasant to watch. If it speaks well and answers well, there is no problem".*

2/ Opinions regarding the social robots in the hospital, advantages, and drawbacks

All participants thought that a social robot could be useful in a hospital.

> *"I think that's where they can be most helpful [at the hospital]. Yes, and in other facilities, where there are a lot of people".*

Some participants noted that it was helpful to make it clear that the robot belongs to the hospital with a distinctive sign to clearly show that it is present to help people:

> *"It's better that it still belongs to the hospital. At the entrance, there are people from the hospital, who are not nurses, but who are paid by the hospital to give us information. You could put a badge on him, something that can distinguish him".*

Participants thought that robots were useful for different tasks:

- Assist hospital staff with practical daily living tasks

  "For care, or helping the nurses when it's busy. In support, yes. The nurse is there to care, not to run".

  "Practical aspects to help. Yes, it could bring the meals, and it could simplify the task of the health professionals".

  "Get a glass of water if you are a patient and really can't move. Like the robots in the kitchen that do that kind of thing. Other activities like that, fetching a book, even holding the book if you can't. It has a role of helping and not bothering the nurse with that kind of thing".

  "I think it's a good idea in some cases. It can be very useful to help with walking, the person feels less dependent. She will have a tool for her. It's good for every activity that needs a tool to replace the human, if there is not enough staff or to do the tasks which are a little painful. But when it's emotional, I prefer animals".

- Distracting the hospitalized patients

  "It depends on whether I'm in bed or not [if I was a patient]. Am I still going to go to the bathroom, or can't I move? I think the more I'm dependent, the more I'll need a robot. It would distract me".

  "Yes, It would bring the newspaper to a person with a visual deficit, and the person would ask the robot: read me what's on the newspaper? Yes, yes, that could be good".

Some people pointed out that the robot could have a listening role (close to that of a health professional):

"Imagine I was given a sedative but I still have pain. With a robot, I could complain, and tell it that I'm still in pain. I think it can be useful if it has a quality of empathy".

"The robot could have empathy, not sympathy, but empathy. Like a psychologist. It might allow me to get it off my chest".

Some participants said that they did not want a robot to take over medical tasks:

"It's to help. It can free up time that would be much more useful for doctors or nurses to spend on other things, and it takes what is basic and automatic. It would be perfect for me in that context. It doesn't replace the cardiologist in the hospital".

"As long as it does not give medical consultations, but otherwise, it can perhaps also give some advice. It can remind you of the sanitary gestures, or that you have to breathe well, provide recommendations".

Most people thought that the robot could not replace human interaction which should be preserved in the hospital:

"One moment, the robot talked about entertaining us in the waiting room. I only see it passing the time, changing the TV channels. But in a hospital where people are in pain, the robot is not suitable for care. There is a real need for real humans".

"It helps the staff, it's going in the direction of the future, unfortunately. But above all, it will relieve the staff of some absurd tasks. But regarding real relations and exchanges, there must be a human. We will never replace humans. The human has something more than a machine. This understanding, the machine will never have it, even if it is perfect".

"I would always prefer human beings. But it always depends on the level of communication you want to put in. A patient at the end of life will want a human being. It will never replace human beings. Whereas someone who has a cold will probably feel good talking to the robot.

3/ Opinions on the data collection and treatment made by robots

Participants' views on the data collection and treatment were varied.

Some people did not want any data to be recorded.

"I don't want it to record any data. Absolutely not".

Some people thought that non-personal data could be recorded:

> *"I don't mind if I ask at 11:50 where the canteen was, I don't mind if it keeps that information for 20 years".*

> *"No, I don't mind, if we give it the information, if we ask it the questions, if they are not sensitive data. The hospital must also respect professional secrecy".*

> *"I refuse to give the number of my health card when I enter the hospital. It doesn't ask us our names. I won't say who I am, I ask to be oriented, to be informed. So for me, it's fine. But obviously, if I add my name, the name of the cardiologist. It's information that I'm seeing the cardiologist and that I have a certain fragility. For me, the robot must not identify me. It should not be nominative".*

Some people thought that confidential data could be recorded by the robot as long as it was treated as sensitive data and remained within the hospital:

> *"It's already done with other devices, so I think we're used to the fact that computers keep information about our lives. After that, why should the interaction with the robot be confidential?"*

> *"It all depends on where data is kept, for what purpose, and how many people have access to the data. You have to put up barriers. So if it's just the hospital [who deals with the robot's data management], no worries".*

> *"I don't mind if it [the robot] has the hospital's agenda and you come and say who you are to go and see such and such a doctor and it answers yes, you are Mrs. what's-her-name, go to 2e floor, consultation room 34, at the end of the corridor".*

> *"If I go to the bathroom 20 times a day, if the insurance company finds out about it, they will lower my policy because [that could mean] I need a prostate procedure. You don't want them to know too much. We need to keep it [the data] within the hospital setting. No more than that".*

Other participants thought that it was important that the robot knew their personal data in order to help them better:

> *"The first thing the robot should ask is who we are. It doesn't know if I'm a man or a woman and it doesn't know what we're here for. Maybe I'm in the wrong hospital. That's the first thing that's missing, it seems to me".*

4/ Concern about the mistakes (bugs, accidents) that social robots might make/cause

Some participants were unconcerned about robots making mistakes:

> *"They are not the only ones who make mistakes. Humans do too".*

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

"I think a machine makes fewer mistakes than a human when it is well programmed".

Most participants thought that robot errors were not serious if they did not involve care:

''It depends on what they [the robots] are assigned to do. If it is about care procedures, it could be a problem".

"If it [the robot] gets the wrong medication, yes, it's serious, but otherwise I don't think so".

"Yes, I am concerned if it [the robot] provides some specific treatment to the wrong person or loses the data".

"As long as it [the robot] doesn't affect the individual, there is no problem".

For some participants, it was essential that the robot could be trusted:

"Yes, totally. We have to be able to trust it [the robot]".

5/ Views regarding the possibility of having an emotional attachment to a robot

People's opinions were diverse. Some people thought that one could have an emotional attachment to a robot. Some were embarrassed by this attachment and others considered it natural.

"Yes. You can get attached to plastic. I realized this the first time I came here. It [the robot] looked at me with its little eyes, it was quite touching. It happened in a few minutes, so if it lasts longer, yes I think it's possible".

''Of course I do think you can have an emotional attachment. People do have emotional bonds with their animals. If I'm lonely, it's very possible" [to be attached to the robot] .

"Attachment yes, it's a presence. But not affection, I won't go that far".

"Yes, but it depends on the robot. I don't mind. No. We get attached to objects, there can be a sentimental value".

"That's what I fear. I've seen people who had empathy for the robot, like with Paro. To have empathy for it [the robot] , that's what I fear".

Other people considered that the robot could be pleasant and useful but that one could not have an emotional attachment to it.

"It's still a doll. To have feelings [for the robot] seems to me a little excessive".

"You don't get attached to it [the robot] like an animal. An animal is a living being. The robot is not alive".

"If it's a patient who is a frequent visitor to the hospital. He may have some confidence in the robot that will direct him right away. It can create some comfort by repetition. It's the robot that is always there, I recognize it. But when it comes to loving it, no. But some form of recognition".

''We have more sympathy, we like to use it if we like it, but it doesn't go any further".

''To me, it's still an object. A hammer is a hammer. And if I run out of hammers I'll take a nutcracker".

''I am not worried. It can fill a gap. Like a food processor, you can use it when you need it, without it missing it when it is not there."

"It [the robot] is an object that will not have empathy".

"No, I find that everything is material, even if there are many things that I like very much [about the robot], each one must remain in its place. There is nothing like human relationships. I can appreciate the robot, but there will be no emotional connection".

6/ Opinions on the risk of the replacement of the human being by robots

Some people thought that machines had already replaced a number of human tasks:

"Robots are already everywhere, without knowing it. The humanoids are another thing, it is perhaps needed to make it accepted by us. But we have already accepted the phones".

"Whether I ask a question to a robot, or on a tablet, I don't see the difference. It already exists, we already have to interact with screens, when we go to a clinic, we register from a screen or with the reception service. It could be a machine like this. So it doesn't matter what the medium is, whether it's a keyboard or voice".

Some people noted that it could be useful for the robot to replace the human in some basic tasks.

"Because I think that it must happen one day or another. I think it can free humans from some repetitive or uninteresting tasks. It can contribute to social progress. But, robots should not dominate the world and replace humans, as in some movies or fiction".

Other people said that robots could not replace humans in relational aspects and worried that robots could gradually replace them.

> *"Robots must have limits. Humans can be replaced for some tasks, for instance, the hard ones the robot can do. The human can stay focused on the relational aspects".*
>
> *"In my job I want him to replace me, but not more".*
>
> *"I hope not. It seems out of place to me. A robot shouldn't be like Blade Runner. They look human and they perform as well. So I think when you're in the hospital, you're in trouble, you're in pain, you really need real empathy. I don't see how the robot could feel anguish or distress".*
>
> *"It bothers me, especially for the emotional. But it still bothers me that in some activities, it's material like robots that are used. Even for sexuality, it already exists, and that bothers me, there is a disturbing side. If it's a tool that helps mobility, yes. But not more. There is a utilitarian side that must remain as such".*
>
> *"Yeah. I'm totally freaking out. It's amazing. What scares me is that the machine dispossesses human beings, that's what scares me".*
>
> *"I'm a little worried that one day I'll be in front of a robot doctor. I don't want robots to replace human beings".*

However, some people thought the robot could replace humans in relational and even intimate tasks and were not bothered by the idea.

> *"Yes, a hospital is a place of distress. We are fragile. So you need support. But right now, there are not enough health professionals, so maybe robots could be useful".*
>
> *"And in some nursing homes, where there is abuse…. Rather than suffer abuse from humans, because unfortunately, nurses break down, could a robot come and bring consolation, and entertainment to people who feel mistreated? [the robot] could bring presence and care. If the nurse is overwhelmed, she can send the robot".*
>
> *"But for example, dirty chores, changing diapers, I would like that from a robot. I would feel less embarrassed if a robot changed my diaper instead of a human. I've been thinking about it for a long time".*

7/ Inferential statistics on the acceptability and usability of the robot for Waves 1, 2, and 3

Following the three evaluations, inferential tests were applied to the Acceptability E-Scale and the System Usability Scale to examine potential significant differences between the 3 evaluation waves.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

The distributions did not follow a normal law and although an analysis of various is robust to assumption violation, It has been decided to conduct non-parametric analysis of variance, a Kruskal-Wallis test for the Acceptability E-Scale and a Mann-Whitney test for the System Usability test.

Regarding the Acceptability E-Scale, a significant statistical difference was observed in regard to Waves 1, 2 and 3 (test de Kruskal-Wallis χ^2=9,83 ; df=2 ; p<0,007). A Dunn post-hoc analysis revealed that in Wave 1, the participants reported significantly higher scores than in Wave 3 (p=0,002). No significant differences were observed between Wave 1 and Wave 2(p=0,067) or between Wave 2 and Wave 3 (p=0,36). Table 10 presents Kruskall-Wallis test results and Table 11 indicates the results of the post hoc pairwise p-values

K (Observed value)	9,83
K (Critical value)	5,99
df	2
p-value	0,007
alpha	0,05

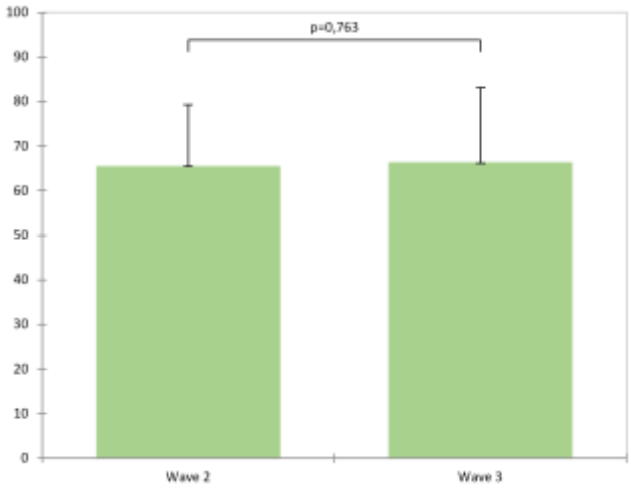
Table 10. Kruskal-Wallis test results

	Wave 1	Wave 2	Wave 3
Wave 1	1	0,067	0,002
Wave 2	0,067	1	0,359
Wave 3	0,002	0,359	1

Table 11. Pairwise comparisons p-values

For the System Usability Scale, a Mann-Whitney test was conducted and no significant statistical differences were found between Wave 2 and Wave 3 (U=188,5 ; p=0,763). The same analysis was conducted after the outliers were withdrawn from both distributions but the results were once again not significant (U=161 ; p=0,297) but a tendency can be observed on a descriptive level between wave 2 and wave 3(m_{Wave2}=68,7 ; m_{Wave3}=72,1). Figure 21 shows the difference between Wave 2 and Wave 3 with the original data and Figure 22 presents the results without outliers.

Due to the small sample size, It has been decided not to conduct sub-analysis with gender, age, or socio-cultural as a between factors but such analysis will be conducted for the final evaluation.



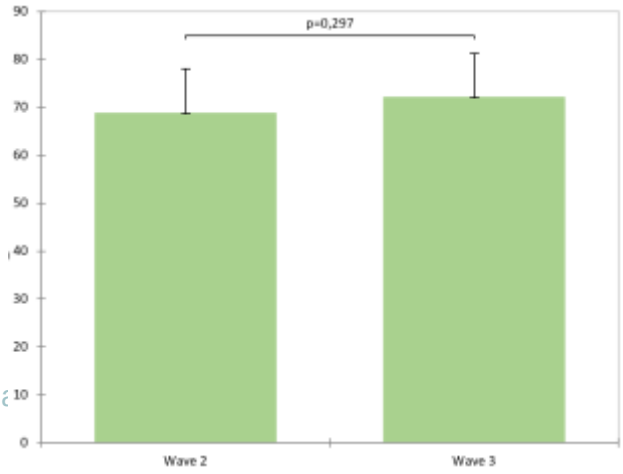


Figure 21. SUS original data from Waves 1 & 2

Figure 22. SUS data from Waves 1 & 2 without outliers

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

4. General discussion and conclusions of the intermediary assessments

The three intermediate evaluations allowed the identification of several arguments explaining the good **acceptability** observed of the SPRING robot in the hospital context. The lack of staff appears to be a major driving force behind the use of socially assistive robots in hospitals. User feedback showed that the robot's usefulness could go beyond information and entertainment which, in itself, was considered a very useful use case if the robot shares clear and accurate information. The robot could assist disabled people to walk or bedridden patients who need help, or who do not wish to "disturb" health professionals for a minor request, such as bringing a glass of water or the TV remote control, for example. In this sense, the robot could help counteract the discomfort of disturbing the staff, and be useful for small tasks. Some participants mentioned that the robot could be useful for maintaining the dignity of the patients. For example, one participant said that in case he would be disabled and bedridden, he would be more comfortable and less embarrassed with a robot doing the changing of diapers than with an aid nurse. These statements are perfectly aligned with the observed quantitative results from the acceptability e-scale as the perceived usefulness was one of the highest scoring items in every wave of evaluation.

The set of evaluations also allowed us to observe varied, and sometimes divergent, positions, among users. For instance, some of them thought that the robot use was not appropriate in a geriatric hospital. It was mentioned that older adults might not be used to technology and robotics. They might also have cognitive disorders or sensory disorders (listening or hearing) and be unable to interact with the robot. The question of which is the user profile best adapted to the use of the robot needs to be better studied. The need to guarantee a high accessibility of the robot and of the services proposed by the robot seemed fundamental for users.

Regarding robot accessibility, we were also able to confirm that many users are willing to use the services offered by a robot but that human service, if available, is preferred by the majority. During the interviews it was mentioned that participants would prefer to ask the nursing staff in case of inquiries, rather than a robot. However, if there were no professionals around, they would ask the robot. The participants also thought that people with serious health conditions would prefer to interact exclusively with human beings, while people who are not very ill, or those accompanying them, would appreciate interacting with the robot to have fun and be entertained. The participants thought that people, who are very frail, would prefer human relationships because they are looking for empathy.

The participants globally thought that the robot was easy to use but they had to get used to it. The robot's voice was pleasant for most of the participants. Still, some of them thought that the voice was too robotic, and lacked emotion. The fact that participants found that they would need time

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

to get used to using the robot might explain, to some extent, the mitigated results observed at the System Usability Scale where results were quite heterogeneous across Its different items.

The robot did not really provide a sense of dialogue with the users. When the participants asked the robot a question, the waiting time for the answer was found to be long. None of the participants noticed the change of color of the robot ears, when it listened, or when it spoke. Many felt that the robot did not understand them, which led to a negative attitude towards the robot, thus impairing the sense of enjoyment during the interactions. The use of subtitles on the robot screen was globally appreciated to facilitate understanding and give confidence to the user, however the service did not always work properly. This caused confusion for some participants. The users perceived many non-adapted answers from the robot, this one not taking into account the previous exchanges or the information the users had already given. Some of the answers, especially the ones regarding orientation within the facility, were too long and difficult to understand orally without visual support on the robot's tablet.

In the evaluations participants appreciated the **humanoid appearance** of the robot, which made it more accessible to interaction with human beings than any other form according to them. Since the main role of this type of "assistive robot" is to give information to the user, participants thought that human attributes could contribute to a better acceptance of the robot. Nevertheless, some participants declared that the robot should not be totally human-like since too much realism in the design may lead to confusion and entail a certain risk of deceiving the potential users of the robot.

Regarding **ethical questions**, participants' opinions were also diverse. Most participants considered that robots in the hospital should be supplementary to care workers and that they should never replace humans. Regarding the **recording and storing of information by the robot**, the vast majority of participants thought that everything depends on the purposes of the use of the data and the way in which the confidentiality of the users' data is respected. Overall, the participants did not express any particular concern about the errors that the robot could make, in the case of using the robot mainly to provide information. On the one hand, they were rather confident in the robot programming (reliability of the computer systems), on the other hand, the participants considered that the potential errors that the robot could make would not have important consequences, apart from being a little annoying. However, they thought that if the robot had to provide care to patients (give treatment), it should not make any errors.

Regarding the potential **attachment to a robot,** the opinions of the participants were quite divergent. For some people, such an attachment was not possible and they did not consider that there was any risk of developing an effective bond with a robot such as the one evaluated. Some people, while declaring that they did not feel concerned by an immediate attachment to the robot, did not exclude that this attachment could develop in the future. For other participants, it could be possible to have a basic emotional bond with a robot, which does not seem to be of concern.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

On the contrary, other participants seemed worried about the attachment they might feel to a social robot. To mitigate this risk, emphasis was placed on the importance of the robot's appearance and behavior being easily identifiable as that of a robot.

Positions on the risk of **human replacement by robots** varied. Most participants acknowledged that such a risk exists and discussed some of its consequences. Some insisted on the importance of defining the role of the robot as complementary to human assistance. Others accepted it in a rather positive way, seeing it as a solution, since the current situation of labor crisis in health care facilities is not going to improve. In a more pragmatic way, these people expressed a degree of acceptance of the replacement of the human by the robot.

Finally, we observed a contrast between the quantitative results which were rather positive with average scores between 3.5/5 and 4.5/5 for most of the scales and the qualitative interviews during which the participants expressed negative opinions, mainly about the usability of the robot. We think that this can be explained by the fact that older participants who took part in the assessments were globally enthusiastic about testing new technological devices, especially new robots. They considered that testing robots is fun and that it shows their openness to the future. They also thought that testing robots is useful for improving technological devices by providing feedback and suggestions to researchers. Therefore during these experiments, they gave rather satisfactory scores because they thought that the work was on-going and that researchers should be encouraged. However when they were interviewed, they clearly expressed their opinions about the flaws of the robot. They also discussed benefits and limits of using robots in the hospital. This emphasizes the need and the benefit of both a qualitative and quantitative analysis to have a comprehensive view of participants' opinions when evaluations of robots in older adults are carried out.

Another interesting point that can be considered as contradictory at first glance is that over the three waves of evaluation, the acceptability scores tend to decrease whereas the usability scores tend to increase. One possible explanation is that in Wave 1, a Wizard of Oz technique has been used whereas in Waves 2 and 3, the robot operated autonomously, making It easier to satisfy the users as a human operator is able to infer thoughts and adapt to the users more efficiently than the robot. Usability was only assessed from Wave 2 onwards as a Wizard of Oz technique has been used in Wave 1 and It is important to note that most of the changes occurred between Wave 1 and Wave 2 (and Wave 3 in a more timid fashion). Technical improvements and newer content have been implemented after each wave of evaluation. Technical improvements can then account for better scores in usability across Wave 2 and Wave 3 whereas the autonomous functioning and the new content (mostly in terms of dialogues, comprehension of what has been said and coherence between what has been said and the given answer) failed to meet the users' expectations. It is then possible to explain why the acceptability scores significantly decreased across the evaluation waves whereas usability scores increased between Wave 2 and 3 as technical improvements have been gradually implemented.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Overall, the results obtained seem fully consistent with respect to the state of maturity of the evaluated solution, which was still under development in the first three waves of assessment of the SPRING project. We confirmed the interest of involving the users in the robot design process, since the suggestions given continuously during the project allow the project consortium to identify ways to improve the device. Our hypothesis is that as the system gains in stability and effectiveness we will find a more positive trend in the evaluated dimensions (acceptability and usability) and less variability in the users' opinions.

The next steps regarding the collection of user feedback will comprise a fourth wave of evaluation in the Living Lab (June-July 2023), which will be followed by a phase of integration of new and improved modules. In parallel and until the end of the project, the SPRING solution will continue to be evaluated in the context of the Broca day hospital with patients and their accompanying persons. These evaluations will allow a wide diversity of users to be included and will provide more solid elements for the evaluation of the relevance of the solution and for its valorization. In terms of valorization of the work done so far, the researchers of the consortium expect to present these results both at specialized conferences and in a scientific publication.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

5. References

Alaiad, A., & Zhou, L. (2014). The determinants of home healthcare robots adoption : An empirical investigation. *International Journal of Medical Informatics*, *83*(11), 825-840. https://doi.org/10.1016/j.ijmedinf.2014.07.003

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, *24*(6), 574-594. https://doi.org/10.1080/10447310802205776

Bevana, N., Kirakowskib, J., & Maissela, J. (1991, September). What is usability. In *Proceedings of the 4th International Conference on HCI* (pp. 1-6).

Brooke, J. (1996, juin 11). *SUS : A « Quick and Dirty » Usability Scale*. Usability Evaluation In Industry; CRC Press. https://doi.org/10.1201/9781498710411-35

Hayotte, M., Thérouanne, P., Gray, L., Corrion, K., & d'Arripe-Longueville, F. (2020). The French eHealth acceptability scale using the unified theory of acceptance and use of technology 2 model: instrument validation study. *Journal of medical Internet research*, *22*(4), e16520.

Micoulaud-Franchi, J.-A., Sauteraud, A., Olive, J., Sagaspe, P., Bioulac, S., & Philip, P. (2016). Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. *Psychiatry Research*, *237*, 196-200. https://doi.org/10.1016/j.psychres.2016.01.043

Peca, A. (2016). Robot Enhanced Therapy for Children with Autism Disorders : Measuring Ethical Acceptability. *IEEE Technology and Society Magazine*, *35*, 54-66. https://doi.org/10.1109/MTS.2016.2554701

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

6. Annexes

Annex 1: System Usability Scale SUS, adapted French version

1. I would like to **use** this robot (talk to, interact with, perform an action) in the hospital as frequently as possible.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

2. I find **conversations** with this robot unnecessarily complex.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

3. I think this robot is **easy to use** (talk, behave, act).

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

4. I think **I will need help** to be able to interact (exchange, talk) with this robot

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

5. I found that the different services (functions) of this robot have been **well thought out (designed)**.

1	2	3	4	5

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

6. I think there are too many **inconsistencies** in this robot.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

7. I imagine that most people would **be able to learn how to use** this robot very quickly.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

8. I found it very **difficult to speak and behave naturally** with this robot.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

9. I felt very **confident** using the robot.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

10. I think there is a lot to **learn to be able to use** this robot well (talk, behave, act).

1	2	3	4	5
Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Annex 2. Acceptability E-Scale AES, adapted version

N°	Items	Scores
1	How easy did you find this robot to use (talk, behave, act)?	
1 – Very difficult **2 –** Difficult **3 –** Moderately easy **4 –** Easy **5 –** Very easy		
2	How understandable were the robot's requests and suggestions?	
1 – Very difficult to understand **2 –** Difficult to understand **3 –** Moderately understandable **4 –** Understandable **5 –** Totally understandable		
3	How much did you enjoy using (talking, behaving, acting) this robot?	
1 – Not at all **2 –** Slightly **3 –** Moderately **4 –** Very **5 –** Enormously		
4	How useful has this robot been to you in (a. greeting you / b. fostering health-safe social interactions / c. helping you prepare for consultations / d. orienting and guiding you / e. entertaining you)?	
1 – Totally useless **2 –** Useless **3 –** Moderately useful **4 –** Useful **5 –** Very useful		
5	Was the time spent by this robot to (a. greet you / b. foster social interactions without health risks / c. help you prepare for consultations / d. orient and guide you / e. entertain you) acceptable?	
1 – Not at all **2 –** Slightly **3 –** Moderately **4 –** Very **5 –** Extremely		

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

6	How would you rate your overall satisfaction with this robot?	
1 – Very unsatisfied **2 –** Unsatisfied **3 –** Neutral **4 –** Satisfied **5 –** Very satisfied		
	TOTAL	…../30

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Annex 3. Observation grid used during the interaction

Dimension	Score (low to high)					
	1	2	3	4	5	NA
1. Understanding of the robot's **speech by users (quality of voice and speech)**						
Observations:						
2. Understanding of **users by the robot (feedback from the transcripts on the screen)**						
Observations:						
3. If the robot understands well (transcripts OK) it **responds in a logical/relevant way**						
Observations:						
4. Understanding of **users' emotions** by the robot						
Observations:						
5. **Behavioral** adaptation of the robot						
Observations:						
6. Ability of the robot to **speak one after another (turn taking)**						
Observations:						
7. Robot navigation and **movements**						
Observations:						
8. **User engagement** during the interaction (attention, involvement, implication)						

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Observations:						
9. **Social proximity of** the person related to the robot (interacting with the robot at a "good" distance)						
Observations:						
10. **Emotional attitude of the user** during the interaction dissatisfied (1) neutral (3) happy (5))						
Observations:						
11. **Remote control** tablet **and functions** (the quality of its functioning)						
Observations:						
12. Global quality of the **human-robot interaction** (Facilitator perspective)						
Observations:						

Particular conditions of the assessment

☐ Person was sitting down
☐ Person was wearing a mask
☐ Person has hearing difficulties
☐ Other (beard, particular voice ?) : :

_____

_____

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Annex 4. Guide for the semi-structured interviews (ethical dimension)

1. Do you **approve** of the development of robots that look like humans? Which are the advantages and the inconvenients ?
2. What do you **think of the use** of social robots in the hospital? Which are the advantages and the inconvenients ?
3. Do you think it's correct/fair (**acceptable**) for robots **to record and store user data**? Why?
4. Do you feel **concerned about the mistakes** (bugs, accidents) that social robots might make/cause? Why?
5. Do you think it is possible to have an **emotional attachment to** a robot? (experiencing feelings, attachment, empathy) Why?
6. Do you think social robots could **replace human professionals** in hospital settings?

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Annex 5. Example Task Sheet

Task 1

You are visiting a daycare hospital. You are at the hospital and this is your first visit. You don't know anything about the building, the schedule of appointments or the amenities available.

Please **talk to the robot and your friend** to try to find the information shown in the image below:

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 871245.

Annex 6. Appreciation of the dialogue capabilities of the robot (HWU questionnaire)

Dimension	Score (low to high)					
	1	2	3	4	5	NA
(1) The robot's voice was clear						
Observations:						
(2) Robot said too much information in one time						
Observations:						
(3) The robot heard me correctly most of the time						
Observations:						
(4) I felt that the robot understood the meaning of my words						
Observations:						
(5) The conversation seemed natural						
Observations:						
(6) The robot heard me correctly most of the time						
Observations:						