The Journal of Supercomputing (2023) 79:17670–17690
https://doi.org/10.1007/s11227-023-05352-7

# ADELA: a conversational virtual assistant to prevent delirium in hospitalized older persons

**Javier Alonso-Mencía[1,2] · Marta Castro-Rodríguez[3] · Beatriz Herrero-Pinilla[3] · Juan M. Alonso-Weber[1] · Leocadio Rodríguez-Mañas[3] · Rodrigo Pérez-Rodríguez[4]**

Accepted: 25 April 2023 / Published online: 9 May 2023


## Abstract

Delirium is a sudden mental state that causes confusion and disorientation, affecting a person's ability to think and remember clearly. Virtual assistants are a promising alternative for non-pharmacological interventions. This research aims to present a prototype of ADELA, a conversational assistant to prevent delirium in hospitalized older persons who speak Spanish. A co-creation process with medical experts to identify requirements was carried out to later develop the assistant iteratively and evaluate it from a technical and usability perspectives; the latest using the Spanish version of the System Usability Scale (SUS) and the Chatbot Usability Questionnaire (CUQ), supplemented by qualitative data. Mean values of 75.5 and 85.94 were obtained for SUS and CUQ, respectively. The technical evaluation helped defining the minimum environmental specifications required for deployment. Obtained results imply the assistant is usable and potentially accepted by the target population; useful information was extracted for refinement. Technical evaluation showed positive results, indicating it can be used in a real clinical environment.

**Keywords** Delirium · Virtual assistant · Conversational assistant · Artificial intelligence · Older persons · Hospitalization · Mental health

✉ Rodrigo Pérez-Rodríguez
rodrigo.perez@urjc.es

1 University Carlos III of Madrid, Leganés, Spain

2 Foundation for Biomedical Research-Getafe University Hospital, Getafe, Spain

3 Getafe University Hospital, Getafe, Spain

4 Intelligent Robotics Lab-Rey Juan Carlos University, Fuenlabrada, Spain

# 1 Background

## 1.1 Clinical context

Delirium is a deterioration in mental functioning that occurs within hours or days and is usually triggered by an acute medical pathology, a trauma, or drugs. It is one of the most frequent medical emergencies, with a prevalence of around 20% in patients admitted to medical services, and even higher in special services such as Orthogeriatric Units or in Intensive Care Units [1]. The prevalence of delirium in hospitalized older patients has skyrocketed 90% in some moments of COVID pandemic [2].

The positive side is that delirium is preventable in 30–40% of the cases [3], being one of the most beneficial strategies for primary prevention multicomponent non-pharmacological interventions [4].

Virtual assistants may provide cognitive stimulation, social interaction, and sensory engagement to older patients during hospital stay. This can be achieved through a variety of methods, such as personalized conversations, games, and activities that are tailored to the patient's interests and abilities. Virtual assistants can also provide reminders for medication and daily routines, as well as information about the hospital environment and procedures to reduce confusion and anxiety [5, 6].

The use of a virtual assistant in the prevention phase of delirium could be particularly useful in contexts where patients are at high risk, such as cases of cognitive impairment, sensory deprivation, or a previous history of delirium [7]. This virtual assistant could also be used in combination with other preventative interventions such as early mobilization, sleep promotion, and medication management, to provide a comprehensive approach.

Smart personal assistants, such as Amazon Alexa, allow searching for information or scheduling events among other functionalities. However, the use of virtual assistants in the health field is still to be explored, so its potential acceptance by older people is uncertain.

The ADELA project, funded by Fundación MAPFRE, aims to design and develop a conversational virtual assistant to prevent delirium in older people (+ 65 years old) during hospitalization. This population is the most prone to develop delirium during hospital admission, due to age-related changes in cognition and physiology [8].

## 1.2 Related work

Balsa et al. [9] proposed an intelligent personal assistant (VASelfCare) to help older people coping with type 2 diabetes. The assistant is integrated in a mobile app and the interaction is done via voice or gestures. A usability evaluation of the assistant obtained a SUS score of 73.75.

Dimeff et al. [10] presented a virtual assistant (Dr. Dave) aimed at reducing unnecessary hospitalizations and suicide events in general patients. To assess usability, authors used the Usability Satisfaction and Acceptability Questionnaire (USAQ), adapted from SUS. A promising average USAQ score of 4.4 (out of 5) was obtained.

Ireland et al. [11] proposed a conversational assistant (Harlie) to treat neurological conditions such as Parkinson's disease. Focus groups were organized to receive feedback from final users, including older people. In general, users had a positive impression, but they found some technical issues and problematic conversational responses.

Inkster et al. [12] developed a conversational assistant (Wysa) embedded in a mobile application focused on mental health for the general population. This system was tested in a study where participants showed depressive symptoms, obtaining an engagement above 70% among the users by doing an analysis through the Patient Health Questionnaire (PHQ-9) [13].

Sun [14] investigated how visual enhanced on voice user interfaces, such as conversational assistants, might mitigate usability challenged by older adults during interaction. The articled concluded that integrating visual output as a feedback mechanism facilitates interaction between older adults and assistants.

Liu et al. [15] examined older adults' preferences for intelligent virtual assistants regarding information modality and feedback. Results showed that the visual-auditory bimodality is superior to single visual modality and single auditory modality for older adults.

Markfeld et al. [16] conducted a study to examine the effect of different feedback modalities (visual and auditory) in a table setting robot assistant for elder care. The visual feedback included the use of LEDs and a screen. The combination of LED lights and verbal commands increased participants' understanding, contributing to the quality of the interaction.

## 2 Objectives

The main objective of this research work is to design and develop ADELA, a conversational virtual assistant to prevent delirium in hospitalized older persons. Specific objectives are:

- To implement a co-creation process with domain experts from the clinical field to extract functionalities and operational requirements.
- To perform an iterative development process along with clinical domain specialists to produce a functional prototype of ADELA.
- To refine the ADELA prototype through a usability study with potential end users. This enhanced version of ADELA will be used in a clinical trial aimed at demonstrating its usefulness to prevent delirium in hospitalized older patients (recruitment to be started in December 2022).

## 3 Material and methods

This section presents the methodology used to conceptualize, design, develop, and refine the conversational assistant.

Figure 1 shows the followed methodology, consisting in 5 phases, to achieve the final version of the conversational assistant.

### 3.1 Co-creation process

The purpose of this phase was to extract information from domain experts to focus and guide the overall design process of ADELA.

The co-creation process was conducted through a series of meetings and workshops, which brought together a heterogeneous group consisting of three medical specialists (geriatricians), that are considered the domain experts, and two technologists. Prior to the meetings, ideas were proposed to reach common ground. The purpose of this process was to enable the development team to obtain a more holistic view of what the system should include and how it should behave. The Delirium Prevention Protocol by the Geriatrics Service of the Getafe University Hospital and the clinical guidelines of the National Institute for Health and Care Excellence [17] were used as baseline to reach the functionalities presented in Table 1.

The Human-Centered Design approach was adopted to ensure that ADELA would be accepted and adopted. This approach involves understanding the users' needs, behaviors, and preferences through research and analysis, and then designing solutions that meet those needs. Participatory design methods, such as design workshops and usability tests, were employed to actively involve users and stakeholders throughout the design process [18].

These methods enabled the development team to gain insights into the users' mental models, expectations, and needs, and to ensure that the design of ADELA aligns with these factors. The use of participatory design is especially important in cases where the adoption and acceptance of new technologies by potential users is low, as it helps to create a usable and successful software system by perfectly adapting the design of the software to the mental model of the potential users [19].

The workshops were used to specify all the relevant aspects of the users and the system's context of use and to translate them into a system's design. These workshops involved focus groups [20] participated by the co-creation team. The second method, usability tests, was later used to evaluate the effectiveness of the first designed prototype in meeting user needs and expectations.

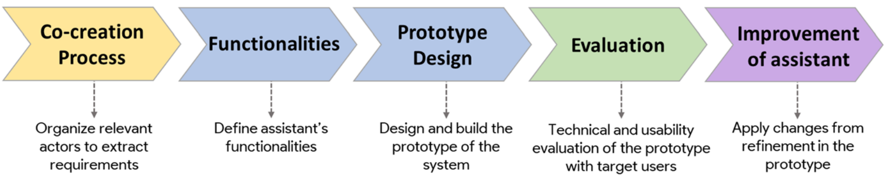


**Fig. 1** Summary of phases involved in the project's methodology

**Table 1** Functional requirements extracted in co-creation phase

| Functional requirement | Clinical justification | Description | Technology demand |
|---|---|---|---|
| Spatial-temporal reminders (time and place where the patient is) | Avoiding disorientation | Periodic reminders to orient the older person about the current time and date, where he/she is and how long he/she has been hospitalized | Integrated speaker assistant |
| Personalized recommendations for physical activity | Prevention of immobility | Personalized physical activity recommendations based on a personalized medical prescription considering physical restrictions, use of catheters, intravenous medication, need for assistance, etc | |
| Personalized recommendations for cognitive activities | Cognitive stimulation | Personalized recommendations for cognitive activity based on a personalized medical prescription | |
| Hydration reminders | Preventing dehydration | Periodic reminders for fluid intake | |
| Reminders to use visual and auditory aids | Avoiding deafferentation | Periodic reminders, if necessary, of the use of visual and auditory aids | |
| Bedtime reminders | Maintaining adequate sleep hygiene | Indications on when it is appropriate to sleep | |
| Soothing sounds | | Playing, if the patient wishes, relaxing sounds to accompany rest | |
| Lighting control | Maintaining circadian rhythms | Automatic adjustment of the brightness of the room according to the cycles of the day | |
| Receiving calls from family members/close contacts | Avoiding isolation | Possibility of receiving external calls through the virtual assistant (at specific time slots) | Assistant integrated in speaker and smart bulb |
| Answering questions | Avoiding disorientation | Conversational capabilities of the assistant to solve doubts (e.g., where he/she is, where his/her family is, date, time, etc.) | Assistant integrated in the speaker (with microphone) |

## 3.2 Functionalities

Given the set of requirements identified in the previous phase, ADELA was designed to have the functionalities described below.

### 3.2.1 Basic intents

ADELA incorporates a set of basic functionalities that are launched by the user, called intents, which were developed to follow the flow shown in Fig. 2.

To activate basic intents, it is necessary for the user to wake the assistant up by using the wake word "Adela". A similar system is used in assistants such as Amazon Alexa [21]. A pre-trained neural model known as Porcupine from PicoVoice platform [22] was used for this task.

Each basic intent has been implemented in AWS Lex, a cloud service capable of extracting a user's intent from a recording. Additional cloud services such as AWS Lambda (to process intent responses) and AWS Polly (text to human speech) were used to complete this functionality.

The assistant incorporates 25 intents in total such as asking for the time, weather information and start a memory game, among others.

### 3.2.2 Reminders

ADELA provides reminders throughout the day, keeping users active with different activities and helping them remember things. Eight types of reminders, described in Table 1, have been included. The list of reminders is stored in an AWS DynamoDB cloud service database.


**Fig. 2** Basic intent workflow in ADELA conversational assistant

### 3.2.3 Playing relaxing music

According to Table 1, ADELA offers the possibility to play relaxing music during the user's preparation for sleep, to create an appropriate atmosphere. To this end, several relaxing songs have been incorporated into the assistant.

### 3.2.4 Phone calls

ADELA can receive phone calls to facilitate communication between patients and their relatives, thanks to the integration of the Simcom SIM868 module [23]. This module works like a mobile phone.

A whitelist with relatives' phone numbers is stored within AWS DynamoDB service to avoid nuisance calls.

### 3.2.5 Lighting control

According to Table 1, ADELA must be able to control the illumination of the room where it is located to help patients maintain circadian rhythms. The lighting control has been implemented using a table lamp incorporating a smart bulb, but it is scalable to more complex settings.

The assistant automatically turns on and off the room lighting throughout the day, but it can also receive commands from the user.

### 3.2.6 Cognitive games

ADELA incorporates six different memory games to promote cognitive stimulation. Games included are explained in detail in Appendix 2. These games require cloud services such as AWS Lex, Dynamo DB, and Google Speech-to-Text.

## 3.3 Prototype design

The design of the working prototype was conducted partially in parallel to the functionalities phase described above.

### 3.3.1 Conceptual architecture

Figure 3 represents the conceptual architecture of the system, including ADELA assistant. Components and relationships are described below:

- The ADELA virtual assistant is the central technological component. The older person communicates with the assistant using the voice.
- The environment automation consists of a Wi-Fi lamp, that is remotely controlled by the assistant.

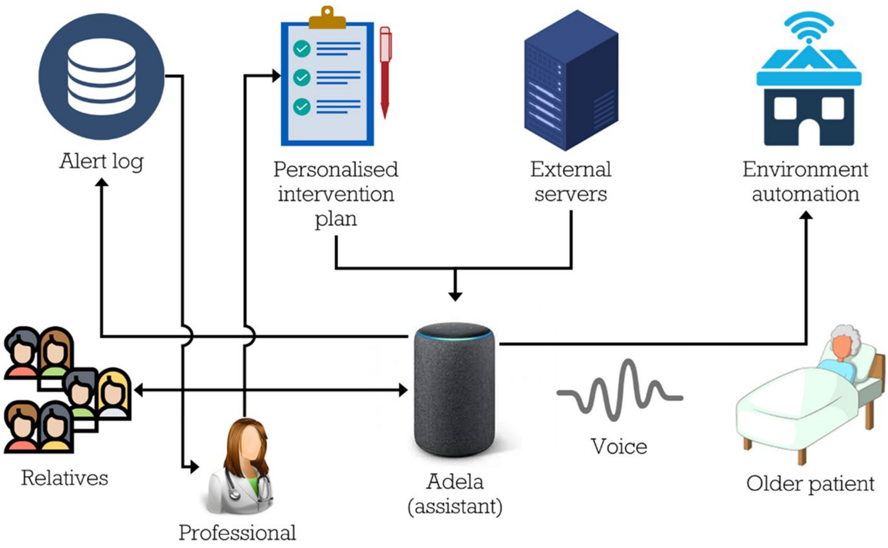


**Fig. 3** Conceptual architecture of the system

- External servers correspond to the cloud services used to meet the assistant's needs, such as a database.
- The medical professionals are responsible for creating personalized intervention plans, and adapting the assistant to the patient's needs and habits.
- An alert log is maintained to be analyzed later in case it is necessary.
- The relatives of the older person can communicate with him/her thanks to the built-in phone call functionality provided by the assistant.

#### 3.3.2 Hardware components

Figure 4 shows the physical prototype of the device running ADELA along with its components. The prototype includes a LED strip, controlled by an Arduino Nano, with different colors to indicate the status of the assistant (see Fig. 5). This functionality has been inspired by the work presented in [14, 15] and [16]. It has been claimed that visual-auditory bimodality in assistants and combining LED lights and verbal commands improves the interaction between assistants and older people.

### 3.4 Evaluation

#### 3.4.1 Usability evaluation

The main objective was to collect feedback and suggestions from potential users to improve the prototype, and thus minimize potential usability problems arising from an improbable design. This is necessary to ensure good usability, user experience and acceptance of the ADELA assistant in a real-life scenario (i.e., first in a clinical trial and later in routine care).

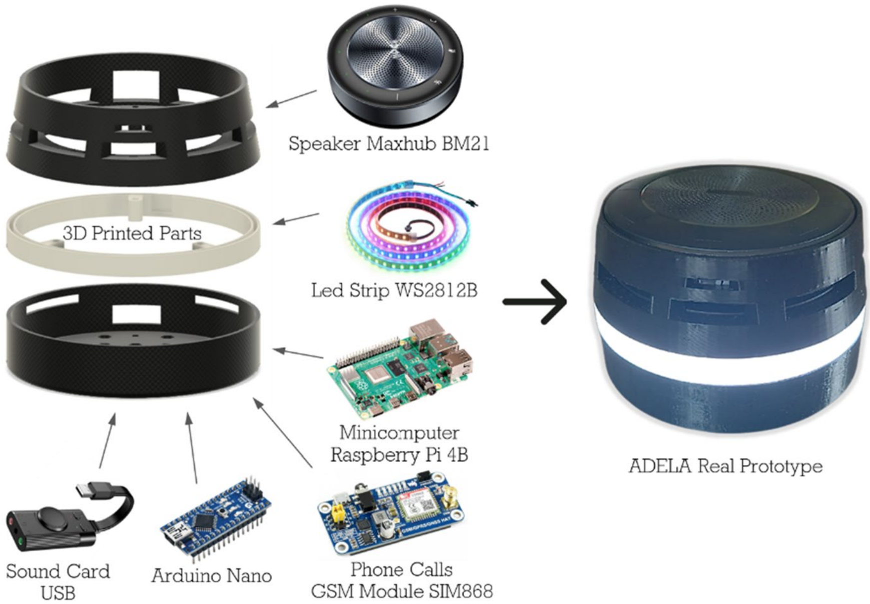


**Fig. 4** 3D model of the assistant together with the components of each section


**Fig. 5** LED colors to indicate different status of assistant

The experiment consisted of an individual 30-min session with each participant following the script described in Appendix 1.

This evaluation was designed to be an observational study in which ten older persons ($N=10$) participated. Participants, recruited at the Geriatrics Service of Getafe University Hospital, interacted with ADELA in a single working session and provided feedback. The tests were carried out in a room within the hospital dependencies replicating the same conditions of a potentially realistic scenario. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Getafe University Hospital (protocol code A06/22 approved 23 June 2022). Participation criteria were:

- Inclusion criteria:
  - o Age > 74 years.
  - o No previous diagnosis of cognitive impairment.
- Exclusion criteria:
  - o Inability of the participant to understand and use the ADELA system.

Usability is defined as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use [24–26]. Used assessment tools were:

- System Usability Scale (SUS) [27]: short 10-item Likert questionnaire that provides a measure of people's subjective perceptions of the usability of a system. These 10 items can be evaluated from '1–fully disagree' to '5–fully-agree'. Total score ranges from 0 to 100.
- Chatbot Usability Questionnaire (CUQ) [28]: Likert questionnaire consisting of 16 items aimed at measuring usability of chatbots.
- Ad-hoc satisfaction questionnaire: customized questionnaire with 5 open-ended questions to collect general user feedback (see Table 2). These questions were designed to complement aspects that are not directly covered by quantitative tools such as SUS and CUQ so additional fallbacks of the evaluated prototype can be identified and thus overcome.

### 3.4.2 Technical evaluation

Every usability evaluation session was audio-recorded to extract technical conclusions and later analyze the performance of the system and understand its limitations. Data used in this analysis comes from the usability phase ($N = 10$).

**Table 2** Ad-hoc questionnaire

| Number | Question |
|---|---|
| 1 | What did you like most about the ADELA virtual assistant? |
| 2 | And what did you like the least? |
| 3 | Did you understand the instructions provided by ADELA? |
| 4 | Did you find it uncomfortable or difficult to interact with ADELA? |
| 5 | What is your general opinion of the ADELA virtual assistant? |

### 3.5 Refinement of the assistant

This phase consisted in introducing the changes to be drawn from the evaluation of the first prototype with potential end-users. This led to a refined version of the assistant that will be deployed in a clinical trial at Getafe University Hospital.

## 4 Results and discussion

Descriptive statistics are used within this manuscript to analyze obtained results. This methodology is a useful tool for summarizing and describing data such as those collected in these studies.

### 4.1 Usability evaluation

Answers provided to the ad-hoc qualitative questions aimed to extract potential improvements, which were used to guide the refinement process. In general, all participants found the assistant practical and enjoyable. The main shortcoming pointed out was the difficulty in waking the assistant up by pronouncing its name ("Adela"). It was usually necessary to repeat the name of the assistant twice or more to wake it up. Regarding explanations and messages provided by the assistant, they were perceived as clear and easy to understand. It was mentioned that the speed of the assistant's speech was sometimes too fast.

According to other studies where the usability of a system is measured by SUS, a system can be considered usable if its SUS score is higher than 68 [27]. In our usability study, an average score of 75 was obtained. Therefore, ADELA assistant could be considered already usable in its initial version.

Figure 6 shows the percentage of responses (positive, neutral, and negative) in five categories from grouped SUS questions. Most categories were positively

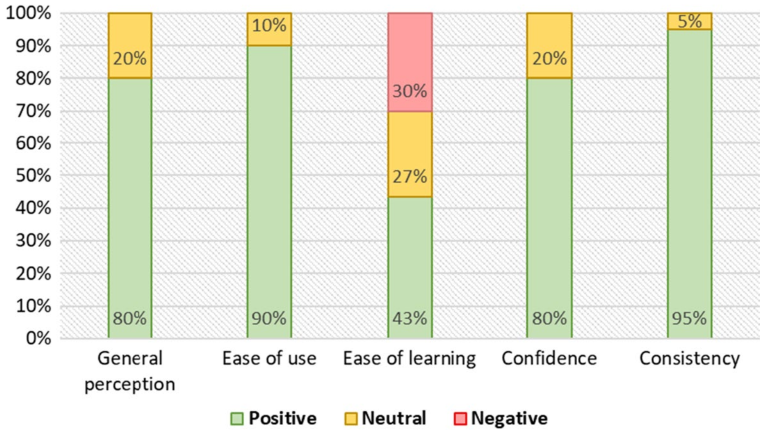


**Fig. 6** SUS answers grouped by categories

evaluated. However, measuring the ease of learning to use ADELA showed that more than 50% of the participants expressed some difficulties.

Regarding CUQ assessment tool, other studies using it [28, 29] also take as a reference a score of 68, from which a system can be considered usable [27, 30, 31]. In this study, an average CUQ score of 85 was obtained, so ADELA assistant can also be treated as very usable.

Figure 7 shows the average score per question in CUQ questionnaire. Odd-numbered questions have a positive meaning if their score is 5. In this case, all odd-numbered questions achieved scores higher than 4, except for question 9, which assesses how well the assistant understands the user. Despite this, participants agree on the assistant's ability to handle and resolve errors (question 13).

On the other hand, even-numbered questions have a positive meaning if their score is 1. The worst results in this group, although positive, are found in questions 8 and 10. Some speeches or interactions from the assistant seems not to be sufficiently clear. Question 16, which indicates whether the assistant is complex, has a score slightly higher than 2, can be considered as an acceptable result, as older people do not see the assistant as complex to use.

## 4.2 Technical evaluation

### 4.2.1 Cloud response time

ADELA assistant uses several cloud services. This means that the system must have a stable and reliable internet connection to provide a feeling of seamless interaction.

The test consisted of asking ADELA a basic intent and measuring the time between the recording was sent until ADELA began to speak. This process was

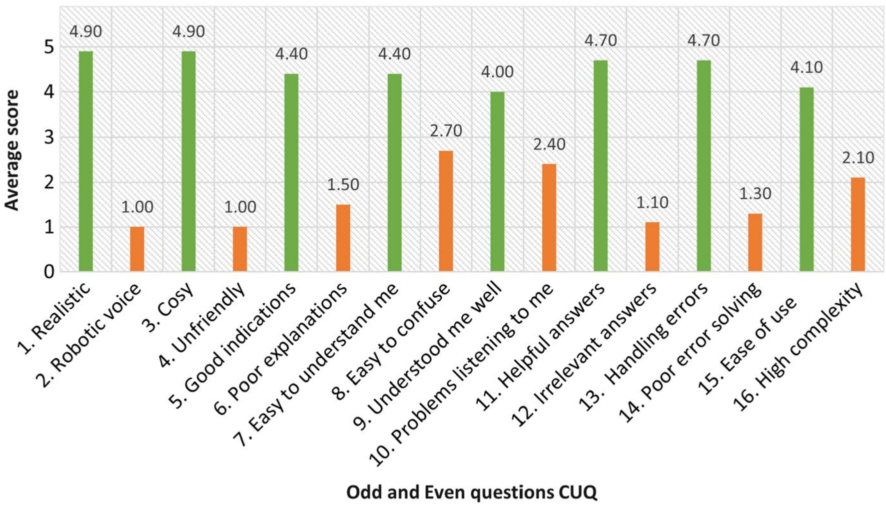


**Fig. 7** Average score on odd and even questions in CUQ. Odd questions score 5 positively, while even questions score 1 positively

**Table 3** Cloud services average response time depending on configuration in ADELA prototype

| Connection type | Internet speed (mbps) | Avg. response time (s) | Standard deviation (s) |
|---|---|---|---|
| Wi-Fi | 2 | 4.97 | 0.49 |
| 4G LTE | 2 | 4.55 | 0.31 |
| Wi-Fi | 10 | 2.63 | 0.05 |
| 4G LTE | 10 | 2.71 | 0.12 |
| Wi-Fi | 100 | 1.90 | 0.02 |
| Ethernet | 100 | 1.87 | 0.03 |

repeated five times with different connection configurations. Results are shown in Table 3.

As mentioned earlier, ADELA will be firstly deployed for a clinical trial at Getafe University Hospital, where the average connection speed varies between 10 and 20 mbps. Average response time will be around 2.7 s, considered appropriate to achieve a reasonable interaction.

#### 4.2.2 Reliability of AWS Lex

AWS Lex is the cloud service used in ADELA assistant to understand users' intent from an audio recording. Therefore, the reliability of this service must be high to assist users.

Intents detection rate from AWS Lex is highly dependent on the way the user speaks, which must be strongly considered when working with older persons.

An evaluation of this service was undertaken to assess its effectiveness in identifying ten different intents with ten potential users of ADELA. Whenever no predefined intent or an erroneous intent was detected, the assistant prompted the user to repeat the question.

The percentage of misguided intents ranged from 0 to 30%. Users 1, 5, 6, 7, 9, and 10 did not exhibit any misguided intents during the observation period. Users 2 and 8 had a relatively low percentage of misguided intents (10%), while users 3 and 4 had 30% and 20% of misguided intents, respectively. The rate of misguided intents is somehow variable among users; an average intent error rate of around 8% was observed. It was also found that all users were able to complete their intents in a second attempt in case AWS Lex did not resolve it the first time. Therefore, in case of understanding errors, the assistant recovers easily.

### 4.3 Refinement of the assistant

The analysis of the results from the refinement phase has led to several ideas on how improve ADELA, which are described in Table 4. These improvements have allowed to build a new version of ADELA ready to be deployed in a real environment.

**Table 4** Problems in ADELA assistant detected in refinement phase with its respective solutions

| Problem | Solution |
|---|---|
| User difficulties in using the ADELA assistant on first contact | Carry out an introductory session where the user is taught on how to use and interact with the assistant |
| The assistant sometimes delivers sentences that are too long and complex | Simplify communication |
| The assistant speaks too fast | Reduce the assistant's speech rate to 85% |
| The detection of the wake word ("Adela") is sometimes not successful | Increase the accuracy of the wake word system by exploring other detection systems approaches or increasing the training examples of the trained neural trained neural model. This task is in progress |
| The user forgets the name of the assistant (Adela) | Place a tag with the name of the assistant (ADELA) on top of the physical device |
| The user forgets that the assistant only listens if the LED lighting turns blue | Provide a short instruction sheet to the user with all the capabilities of the assistant to facilitate learning |
| The user may get confused during interaction in games | Simplify interaction in games and clearly indicate when it is the user's turn to speak |
| Provided instruction for the games are sometimes hard to understand | Change the strategy to explain the game with different and additional explanations |
| The time during which the assistant is listening is too short | Increase listening time by 15% compared to the original |

## 5 Conclusions

ADELA is a research project aimed at designing and developing an intelligent conversational assistant to prevent delirium in hospitalized older people. To achieve this goal, the project went through several phases until a refined system, ready to be used in a clinical environment, was released.

ADELA was built through a co-creation phase in which a multidisciplinary team elicited the requirements of the assistant. Thanks to this co-creation phase, a first functional prototype of the conversational assistant was designed and built. This first prototype was exposed to a technical and usability evaluation process conducted with 10 potential end users (older adults). SUS and CUQ were used with a mean score of 75.5 and 85.94, respectively.

Changes to improve the first ADELA design included, among others, simplifying voice interaction and provide the user with simple information on how to use the system. The technical evaluation allowed identifying a minimum requirement of a 10mbps connection in the environment where the assistant is deployed.

It is important to acknowledge some limitations of this study that may have affected the interpretation of the results. First, the potential bias introduced by the population participating in the study; it is possible that those who volunteered had a higher interest or motivation to participate than others, which may have influenced the results. Additionally, it is important to note that the study was conducted with healthy older individuals, and not with those who would typically use the system in a real-world context (older persons with a condition leading to being hospitalized), which could limit the generalizability of the findings. From a usability standpoint, these limitations may result in a positive bias towards the system.

Furthermore, it should be noted that some technical aspects could also present several limitations. For example, the reliability of Lex was tested with a relatively small sample size of 10 individuals, which may not be sufficient to fully evaluate its reliability in different contexts. Moreover, due to the variety of voices, accents, and speech problems, the results may not be conclusive and may require further investigation. These limitations should be considered when interpreting the results of this study.

Finally, a second version of the assistant, ready to be deployed in a real environment, was released including the improvements drawn from the evaluation phase.

Future work will comprise deploying ADELA in a real scenario to assess its clinical validity in preventing delirium throughout a clinical trial. Furthermore, new features will be added to the system to enhance its behavior. Among these functionalities is the inclusion of Natural Language Processing (NLP) with models such as GPT-3 to expand the assistant's conversation options, reduce its cloud dependency by using offline services and explore new wake words detection methods to improve user experience.

## Appendix 1: Script for ADELA's refinement

*Approximate duration* 30 min

Before starting the activity, the elderly user will be informed about the following aspects of the ADELA assistant:

- How the assistant should be woken up.
- The assistant can be asked questions both 'waking her up' and when she initiates the conversation.
- Explanation of the physical exercise sheet.
- The assistant can receive calls.
- The assistant can be asked to play cognitive (or memory) games.
- Color coding of the LEDs on the device according to their status.

The activities to be carried out in the test, divided into sections, are set out below:

*Section 1* Check basic interaction to activate the wizard and ask for the current time.

- Objective: To test basic user-assistant interaction, as well as the phrase used by the user to request the functionality.
- Technical details: Not applicable.
- Relevant information: The assistant is called 'Adela' and will be activated when you tell her your name.
- Activities requested from the user:

Request time/day.

- Expected result: The assistant will tell you the current day and time.

Know where I am.

- Expected outcome: The assistant will tell you where you are.

Know where your relatives are.

- Expected outcome: The assistant will tell you the location of your relatives.

Ask the assistant who he/she is and what he/she does.

- Expected outcome: The assistant briefly describes himself/herself and explains his/her capabilities.

Ask the assistant to turn up the volume.

- Expected outcome: The assistant turns up the volume.

Ask the assistant to turn down the volume.

- Expected result: The assistant turns down the volume.

Ask the assistant to play soothing music.

- Expected result: The attendant will play music.

Ask the attendant to stop the music.

- Expected result: The attendant will stop the music.

Ask the attendant to turn on the light.

- Expected result: The attendant will turn on the light.

Ask the attendant to turn off the light.

- Expected result: The attendant will turn off the light.

*Section 2* Full flow in hydration reminder.

- Objective: To determine the time required by the user to complete flow.
- Technical details: It will be forced in real time to launch the reminder.
- Relevant information: The user will be provided with a table with a bottle of water. The wizard will prompt the user with a recommendation for water intake.
- Activities requested from the user:

  Follow the instructions of the wizard.

  - Expected result: The wizard triggers the hydration reminder, and the user takes time to drink water.

*Section 3* Full flow in reminder in reminder visual/auditory aid.

- Objective: To determine the user's response time.
- Technical details: The reminder will be forced in real time.
- Relevant information: The user will be provided with a table where he/she should leave his/her glasses/hearing aids (if used). The assistant will indicate the user to put on the visual and hearing aid.
- Activities requested of the user:

  Follow the instructions of the wizard.

  - Expected outcome: The assistant launches the visual/auditory aid reminder, and the user takes time to put on their device. The assistant waits for 30 s to listen in case the user needs anything else.

*Section 4* Full flow and time required per phase in physical activity performance.

- Objective: To determine the user's response time.
- Technical details: The reminder shall be forced in real time. The physical activity plan will be limited to 1 exercise.
- Relevant information: The user will have an exercise sheet on a table.
- Activities requested from the user:

  Follow the instructions of the wizard.

  - Expected outcome: The assistant instructs the user to pick up the exercise sheets from his or her desk and waits 10 s. The assistant then tells the user the partial physical plan to perform. After 2 min (20 min for a complete plan), the assistant will ask the user if they have finished their activity and will say goodbye.

*Section 5* Testing user-assistant interaction in the phone call function.

- Objective: To determine whether the sound quality and call acceptance and rejection mechanisms are effective.
- Technical details: An evaluator's phone number will be added to the database to simulate a call with a family member.
- Relevant information: When receiving a call, the assistant will ask whether to take or reject the call. At the end of the conversation, the user cannot hang up the call. The person who made the call to the attendant must hang up the call.
- Activities requested from the user:

  The user picks up the call.

  - Expected result: The attendant receives a call, and the user confirms that he/she wants to take the call. During the call both the sender and the receiver can communicate clearly and have a conversation. When the conversation is completed, the sender hangs up the call to end the process.

  The user rejects the call.

- Expected result: The attendant receives a call, and the user rejects the call. The call is terminated, and the normal flow of the attendant continues.

*Section 6* Evaluate flow of each cognitive game.

- Objective: To determine the level of understanding and difficulty of each game for the user.
- Technical details: In the last game, it will be forced in real time to launch the reminder.
- Relevant information: The user must activate the assistant by her name Adela and ask the assistant to perform a cognitive activity.
- Activities requested to the user:

Number addition game.

- Expected result: The user performs the addition of 3 numbers indicated by the assistant and the assistant indicates if the result is correct.

- Number counting game.

- Expected result: The user counts the numbers in the indicated interval for 20 s and the assistant indicates the success rate.

- Memory game.

- Expected result: The user correctly repeats the sequence of words spoken by the assistant during the 5 rounds of words corresponding to the game.

- Classification game in categories.

- Expected result: The user indicates the category to which an object belongs from the possibilities indicated by the assistant, and the assistant corrects his/her answer.

- Attention game.

- Expected result: The user counts the number of times a specific word appears in a list of words given by the assistant, and the assistant corrects the answer.

- Random game (triggered by the daily reminder to perform cognitive activity).

- Expected outcome: The user has to decide the appropriate word among several options with respect to the context of a sentence of the assistant, and the assistant corrects his answer.

*Section 7* Checking the flow at bedtime.

- Objective: To determine the user's ability to respond to requests from the wizard and to time the flow subparts.
- Technical details: The corresponding reminder will be advanced in the test.
- Relevant information: You will be asked if you want to reduce the lighting in the room as well as play relaxing music during the process.
- Activities requested from the user:

  Rest reminder.

  - Expected outcome: The assistant indicates that it is time to get ready for bed. The user is asked if he/she wants to lower the lighting by 50% and play relaxing music. After 2 min (10 min in real environment), the lighting will be switched off and the music will be stopped, the user is asked if he/she is ready to go to sleep and the system will be switched off.

## Appendix 2: Memory games included in ADELA

1. *Adding numbers* The assistant prompts the user with several numbers to perform the addition (e.g., *What is the result of adding the numbers 10 and 5?)*
2. *Counting numbers* The assistant prompts the user for an interval between two numbers, as well as the step between numbers to complete the sequence. (e.g., *You must count from 70 to 80 in steps of 2*.)
3. *Word repetition* The assistant prompts the user for a sequence of words to repeat. The length of the sequence shall be increased by one more word (up to a maximum of 5) each time the user gets the previous sequence correct. (e.g., *[Dog], your turn; {User response} … [Cat Dog], your turn; {User response} …*)
4. *Category classification* The assistant prompts the user with a word that he/she must classify into a category from a set of options. (e.g., *Is a rose an animal, a piece of furniture or a flower?)*
5. *Attention* The assistant specifies a word that the user must count in a series of uttered words. (e.g., *I will tell you different animals and you must count the number of times I say the word, Cat. [Dog, Elephant, Cat, Dog, …]*)
6. *Right word* The assistant will display a definition along with several possible options and the user must select the appropriate word. (e.g., *A metal object that fits into locks and allows doors to be opened. Possible options are keys or books.*)

**Author contributions** RPR is the ADELA project's Principal Investigator, he has coordinated the whole process, including this manuscript; himself, along with MCR conceptualized the virtual assistant and obtained funds to design, develop and validate it. JAM did most of the technical work, under the supervision of RPR and JMAW; he also led the writing process of the paper, requesting contributions from all authors. MCR, BHP and LRM led and supervised all clinical aspects of the conversational assistant. All authors reviewed the manuscript.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was supported by Fundación MAPFRE. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

**Data availability** The data that support the findings of this study are available upon reasonable request from the corresponding author, RPR. The data are not publicly available due to privacy of research participants.

## Declarations

**Conflict of interest** Authors have no competing, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Getafe University Hospital (protocol code A06/22 approved 23 June 2022.

**Informed consent** Informed consent was obtained from all individual participants included in the study.



## References

1. Ryan DJ et al (2013) Delirium in an adult acute hospital population: predictors, prevalence and detection. BMJ Open 3(1):e001772
2. Wong EK-C et al (2022) Characteristics, treatment and delirium incidence of older adults hospitalized with COVID-19: a multicentre retrospective cohort study. Can Med Assoc Open Access J 10(3):E692–E701
3. Rohatgi N et al (2019) Initiative for prevention and early identification of delirium in medical-surgical units: lessons learned in the past five years. Am J Med 132(12):1421–1430
4. Ludolph P et al (2020) Non-pharmacologic multicomponent interventions preventing delirium in hospitalized people. J Am Geriatr Soc 68(8):1864–1871
5. Inouye SK, Bogardus ST Jr, Baker DI, Leo-Summers L, Cooney LM Jr (2000) The Hospital elder life program: a model of care to prevent cognitive and functional decline in older hospitalized patients. Hospital elder life program. J Am Geriatr Soc 48(12):1697–1706
6. Huang H-Y, Bashirpour M, Agarwal P, Demiris G (2018) Design of a virtual companion for older adults to support aging in place. J Med Syst 42(9):163
7. Krogseth M, Wyller TB, Engedal K, Juliebø V (2011) Delirium is an important predictor of incident dementia among elderly hip fracture patients. Dement Geriatr Cogn Disord 31(1):63–70
8. Noimark D (2009) Predicting the onset of delirium in the post-operative patient. Age Ageing 38(4):368–373
9. Balsa J et al (2020) Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. J Med Syst 44(7):1–12
10. Dimeff LA et al (2020) A novel engagement of suicidality in the emergency department: Virtual Collaborative Assessment and Management of Suicidality. Gen Hosp Psychiatry 63:119–126

11. Ireland D et al (2016) Hello Harlie: enabling speech monitoring through chat-bot conversations. Stud Health Technol Inform 227:55–60
12. Inkster B, Sarda S, Subramanian V (2018) An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 6(11):e12106
13. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B (2016) Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. BJPsych Open 2(2):127–138
14. N. Sun (2020) CareHub: smart screen VUI and home appliances control for older adults. In: The 22nd International ACM SIGACCESS Conference on Computers and Accessibility. 1–4
15. Liu N, Pu Q, Shi Y, Zhang S, Qiu L (2022) Older adults' interaction with intelligent virtual assistants: the role of information modality and feedback. Int J Hum Comput Interact 39(5):1–22
16. Markfeld N, Olatunji S, Gutman D, Givati S, Sarne-Fleischmann V, Edan Y (2021) Feedback modalities for a table setting robot assistant for elder care. arXiv preprint arXiv:2103.08428
17. Lugo GFO, et al (2022) Incidencia del S\'\indrome Confusional (delirio) en un Hospital Universitario Latinoamericano. Rev Colomb Psiquiatr
18. International Organization of Standardization (2019) ISO 9241-210:2019. Ergonomics of human-system interaction–Part 210: human-centered design for interactive systems. [Online] Available: https://www.iso.org/standard/77520.html
19. Trischler J, Pervan SJ, Kelly SJ, Scott DR (2018) The value of codesign: The effect of customer involvement in service design teams. J Serv Res 21(1):75–100
20. Krueger RA (2014) Focus groups: A practical guide for applied research. Sage publications, California
21. Amazon. Amazon Alexa Official Site: What is Alexa?" [Online]. Available: https://developer.amazon.com/es-ES/alexa
22. Picovoice-Wake Word Detection; keyword spotting. Picovoice Wordmark. [Online] Available: https://picovoice.ai/platform/porcupine/
23. E. (http://www.szmynet.com/). Simcom SIM868. SIMCom. [Online] Available: https://www.simcom.com/product/SIM868.html
24. International Organization for Standardization. ISO International Standard ISO 9241-11. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en. Accessed 10 Feb 2023
25. Brooke J (1996) SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry 189(194): 4–7
26. Brooke J (2013) SUS: a retrospective. J Usability Stud 8(2):29–40
27. Sauro J (2011) Measuring usability with the system usability scale (SUS)
28. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, McTear M (2019) Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces? In: Proceedings of the 31st European Conference on Cognitive Ergonomics 207–214
29. Larbi D, Denecke K, Gabarron E (2022) Usability testing of a social media chatbot for increasing physical activity behavior. J Pers Med 12(5):828
30. Lewis JR, Sauro J (2018) Item benchmarks for the system usability scale. J Usability Stud 13(3):158–167
31. Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 4(3):114–123