

Article

A Context-Aware Conversational Agent in the Rehabilitation Domain

Thanassis Mavropoulos ^{1,*}, Georgios Meditskos ^{1,*}, Spyridon Symeonidis ¹, Eleni Kamateri ¹, Maria Rousi ¹, Dimitris Tzimikas ², Lefteris Papageorgiou ², Christos Eleftheriadis ³, George Adamopoulos ⁴, Stefanos Vrochidis ¹ and Ioannis Kompatsiaris ¹ 

¹ Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece; spyridons@iti.gr (S.S.); ekamater@iti.gr (E.K.); mariarousi@iti.gr (M.R.); stefanos@iti.gr (S.V.); ikom@iti.gr (I.K.)

² Entranet Ltd., 54249 Thessaloniki, Greece; tzimikas@entranet.gr (D.T.); papageorgiou@entranet.gr (L.P.)

³ ELBIS P.C., 57001 Thessaloniki, Greece; ce@elbis.gr

⁴ Evexia S.A., 63080 Thessaloniki, Greece; gm@evexia.gr

* Correspondence: mavrathan@iti.gr (T.M.); gmeditsk@iti.gr (G.M.)

Received: 30 September 2019; Accepted: 30 October 2019; Published: 1 November 2019



Abstract: Conversational agents are reshaping our communication environment and have the potential to inform and persuade in new and effective ways. In this paper, we present the underlying technologies and the theoretical background behind a health-care platform dedicated to supporting medical stuff and individuals with movement disabilities and to providing advanced monitoring functionalities in hospital and home surroundings. The framework implements an intelligent combination of two research areas: (1) sensor- and camera-based monitoring to collect, analyse, and interpret people behaviour and (2) natural machine–human interaction through an apprehensive virtual assistant benefiting ailing patients. In addition, the framework serves as an important assistant to caregivers and clinical experts to obtain information about the patients in an intuitive manner. The proposed approach capitalises on latest breakthroughs in computer vision, sensor management, speech recognition, natural language processing, knowledge representation, dialogue management, semantic reasoning, and speech synthesis, combining medical expertise and patient history.

Keywords: wireless sensor networks; ontology-based dialogue management; semantic disambiguation; natural language processing; virtual assistant

1. Introduction

Health care has always been a person-centric affair, providing personalised services to people in need, leveraging along the way whatever means are available at the time [1]. Technology has been amongst the first coefficients that substantially helped the field evolve into what it is today: An amalgam of medical knowledge, professional diagnostic expertise, and exploitation of latest advancements like the Internet of Things (IoT), smart portable devices, and ambient sensors [2,3]. Behind these technological breakthroughs, there are underlying causal effects like a new lithography process making devices smaller and more energy efficient, an abundance of data offering anonymised and privacy-conscious training material, and a new generation of computational algorithms enabling these devices to process said data faster even on low power devices. Rehabilitation, being a specific section of health care that caters not only to the physical recuperation of patients but also to their mental prowess, ameliorates people’s quality of life, while directly affects society as a whole; the faster injured individuals return to their everyday lives, the better for cutting down on the medical cost, benefiting both national health systems and families while simultaneously maintaining family cohesion.

Assisted-living technologies in rehabilitation, particularly when fused with sophisticated dialogue systems and smart interfaces, have the potential to support hospitals and residences alike in a variety of cases; indicative uses include the exploitation of sensors of different modalities to track user behaviour and to log respective biometric readings (systolic, diastolic, heart rate, and blood pressure measurements), while dialogue-based virtual assistants enhance user experience using notifications for reminding drug dosage, recuperation exercises, etc. When attempting to apply more than one of these technologies to real-life conditions, either in home or hospital environments, many hurdles need to be overcome in order to explore their potential to the fullest. Therefore, contemporary solutions directed to people with motor or mental conditions are currently addressing users' needs only partially, while a more complete offering has not yet been materialised.

The intent of the presented, still ongoing, work is to rectify this deficiency by offering a novel framework that combines (i) natural machine–human interaction based on a dialogue-driven, smart virtual assistant and (ii) a semantic-infused framework capable of exploiting information originating from sources of different modalities (sensor data, camera data, medical knowledge, or patient history records) to conduct relevant analyses and to extract meaningful conclusions. Beneficiaries of the platform would be both home and hospital settings, where it is essential to constantly monitor and support patients in recuperation, a process that would profit enormously by any degree of automation. Thus, the framework, which is being developed in the context of the research project REA (<https://rea-project.gr/en/home-en/>), constitutes an original fusion of distinct components that, in unison, offer advanced functionalities; firstly, speech recognition is conducted, user queries are transcribed to text, and the content is analysed for noteworthy contextual details. The analysis along with information retrieved from sensor readings and camera feed examination are consequently forwarded to the main system to update the internal status of the dialogue manager. This is the system that keeps track of the discussion while also consults the user's medical history in search of an appropriate answer to address the user's query. An ontological framework, through the innate reasoning capabilities that it possesses, is responsible for conducting logical and fluid machine–human conversations. Furthermore, one of the platform's significant features is the capacity of timely notifications to nursing staff concerning probable patient life-threatening situations. The fusion of the aforementioned focal points composes REA's contribution and differentiating factor with regards to other contemporary frameworks that mostly concentrate on specific functionalities. It is encapsulated in the following:

- A sophisticated dialogue-management module is being developed using a novel framework which leverages the advantages of two different categories of solutions: The task-oriented and non-task-oriented conversational systems. Thus, in addition to being able to carry out specific tasks, it also possesses the capability to produce “chatty” style dialogues. These functionalities increase the system's naturalness and user engagement.
- The platform accommodates data from heterogeneous multimodal IoT sensors. Fusing, analysing, and combining data from wearables, cameras, sleep sensors, and blood sugar/pressure levels to monitor the behavior and condition of individuals has not been implemented yet, to the best of our knowledge.
- Likewise, supporting dialogue management from a semantic scope is a field which has not been exploited enough yet in literature. Semantics offer a smart and rich interconnection of the metadata, which combined with the reasoning framework can contribute in the topic selection issue of the dialogue-management system.
- The supported users are not limited to either patients or medical staff, as there have been provisions to cater to the needs of both groups by offering highly sought-after services (based on user requirements).

Rehabilitation centres, hospitals, and clinics form the most probable integration objectives for the presented work, since services like almost real-time person-counting, activity-monitoring, fall-detection, and sensor-driven alert systems contribute to high-priority, patient-centric decision

making. A voice-activated, intelligent, virtual assistant capable of acting like an intermediary between caregiver and patient is another feature that could benefit equally the clinical and home environments; ameliorating patient recuperation time, via the monitoring of sleep quality, medication/diet, or exercise schedule, ambient sensor remote management (light/temperature) enhances the sense of self-sustainability, while also achieves euphoria and mental health. Likewise, caregivers avoid trivial, time-consuming activities (i.e., opening/closing room curtains/TV or altering room temperature) while optimise their routine and focus on important health-related tasks. To achieve this goal, the REA platform needs to satisfy the following criteria: (1) to deploy a robust infrastructure that can accommodate multi-sensor, camera, and voice data; (2) to monitor patient condition and behaviour via the analysis of the abovementioned data; (3) to fuse analysis of data with both medical and behavioural patient background, clinical expertise, and conversational history in order to provide accurate suggestions; and (4) to deploy an interactive virtual agent, capable of supporting all stakeholders through dialogue-based interventions.

Therefore, in order to provide a detailed description of the approach that we adopted to handle the respective research challenges, we organised the manuscript into nine discrete sections. An overview of the state-of-the-art is analysed in Section 2, while in Section 3, there is an analysis of the system's architecture. Section 4 is dedicated to the integrated sensors, the data collection framework, and the data protection precautions that were considered and adopted. In Section 5, the language-understanding and speech-synthesis layers are detailed, while in Section 6, we elaborate on the knowledge representation and reasoning frameworks. Section 7 addresses the challenges in decision making and web-based information retrieval, while Section 8 illustrates the framework's functionality in a real-world use-case scenario. Section 9 presents the conclusions and future aspirations for our work.

2. Related Work

Virtual agents roots are traced to legacy systems which used speech recognition to offer ground-breaking technologies for their time. VAL (Voice Activated Link) and Sphinx-II are such examples of systems that were ahead of their era and were later replaced by more advanced and voice-enabled with dialogue capabilities solutions [4]. Conversational agent's audience has been expanded ever since, and they have been proposed as a means to promote mental well-being [5] or even as an automated social skills trainer for people with autism spectrum disorders [6]. Modern agents encountered in everyday consumer products (Google's Assistant, Amazon's Alexa, Apple's Siri, and Microsoft's Cortana) are mostly an expected feature rather than an exception, and the practice is rapidly expanding to the health-care sector [7], where they are meant to be utilised as portals for medical services. Amazon has enjoyed a head start in providing virtual assistant services, with Alexa being a top-charting device sales-wise, already installed in millions of houses. In addition, outside of home settings, there have been numerous clinical environments embracing the technology recently, conducting pilot testing, and assessing its usefulness [8]. However, Google, Apple, and Microsoft plan to also integrate health-care-oriented services into their respective devices, with Google Home being on the brink of hospital trials as well. Moreover, specialised, medically focused products are being developed, assisting in (i) health-care organisation assistance (Merit (<https://merit.ai/>)) by providing automated appointment re/scheduling and sending outbound reminders; (ii) doctor support (Suki (<https://www.suki.ai/>)), Robin (<https://www.robinhealthcare.com/>) via the implementation of voice commands and generation of accurate notes; and (iii) patient and caregiver support alike (Aiva <https://aivahealth.com>, based on the amazon echo platform) by offering a direct channel of communication between the two groups in the form of a multipurpose mobile app.

In Reference [9], the authors examine under which conditions virtual assistants can provide support to elderly people and people with cognitive disabilities so they can live autonomously. The conversational importance of virtual assistants has been reported in Reference [10], where emphasis is placed on the role of chatbots beyond patient monitoring; it focuses on patient–doctor interaction

over a chatbot-driven telemedicine platform after the former has been discharged from a clinical environment. Finally, a comprehensive review on existing conversational agents for health-related purposes can be found in Reference [11]; the characteristics, current applications, and evaluation measures are laid bare regarding agents with unconstrained natural language-input capabilities. Nevertheless, certain concerns have been expressed over potential safety risks when patients are using virtual assistants as the sole source for medical information [12].

Most dialogue-driven platforms share architectural common grounds, embracing a standardised approach comprised of components like a speech-to-text and text analysis layer, a dialogue manager one, a reasoning mechanism, and a language generation (LG) layer. Speech-to-text manages the transcription of any vocal user input addressed to the system. Robust speech recognition software ascertain that high-quality textual input reaches the text analysis process, tackled by powerful natural language processing (NLP) algorithms. Next, the input is mapped through natural language understanding (NLU) techniques to meaning representations; semantic parsers and machine-learning models are used for the extraction of sentence information such as named entities, speech act, or significant keywords. Discourse analysis results are then handled by the dialogue manager, which is also responsible for other tasks, like manipulating knowledge database queries, using previous analyses and dialogue history to predict subsequent actions, or providing the appropriate answer to user questions. This error-proof input interpretation is achieved by employing reasoning techniques capable of exploiting all available multimodal data, while semantic-based procedures interlink the latter to accomplish optimal situational awareness. In order for the system to communicate naturally its decision to the user, language generation is employed, producing articulate utterances. In what follows, we expand on the most important works of each respective research field.

Nowadays, state-of-the-art speech recognition Application Programming Interfaces (APIs) are commonplace and are being offered by global industry conglomerates (e.g., Microsoft, Amazon, and Google), integrating easily into custom frameworks. Language generation is the process of producing meaningful text from nonlinguistic knowledge representations [13] and has preoccupied researchers since the 2000s, with most contemporary studies (References [14,15]) presenting promising results by utilising Deep Neural Network (DNN)-based systems. DNNs [16] are also being exploited in speech-generation systems, which have attracted much attention as well, with other methods varying between Hidden Markov Model (HMM)-based models [17] and incremental dialogue systems [18]. The popularity of DNNs further expands to text-to-speech efforts, with models that are based on waveforms [19] or on spectrograms [20]. Innovative uses of Long short-term memory recurrent neural networks (LSTM-RNNs) have been explored to various speech applications as can be seen in Reference [21], where the proposed architecture can synthesise natural sounding speech without requiring utterance-level batch processing. Likewise, an LSTM's performance is enhanced via hierarchical cascading in Reference [22].

Aiming for a complete solution for people in rehabilitation, REA was originally conceived as an all-around monitoring tool; to meaningfully affect patient's lives, crucial functionalities need to be supported to improve their day-to-day well-being. The objective can be achieved by leveraging sensor data and analysing aspects like the person's sleep quality or posture. The latter directly affects the quality of life of the elderly and of people with motor disabilities, since reliable posture analysis could lead to trustworthy fall-detection results. Fall-detection systems need to overcome significant challenges to produce accurate results, since a multitude of variations need to be accounted for before a notification is sent out. This is easily evident when accounting the fact that, despite various efforts, till now, no platform has managed to reliably address all issues; most popular research directions rely on ambient sensors [23], gyroscopic sensors and accelerometers [24], and computer vision techniques [25]. Furthermore, popular directions that have been exploited to monitor a person's quality of sleep, an issue with broad ramifications and immediate effects on his/her well-being according to relevant studies [26], rely on users' self-reports or entail the connection of dedicated sensors throughout the duration of sleep [27].

An essential unit of any dialogue system is the module that manages the interaction between the agent and the user. Several solutions have been suggested for the implementation of the dialogue manager. All the proposed implementations conduct the tasks of a) representing the last user utterance in a format which is named a dialogue state and b) deciding on the most suitable system action to be executed at each dialogue turn [28]. Handcrafted, ruled-based, and statistical approaches have been proposed in the literature for the procedures of representing/Updating the dialogue state and formulating a reasonable system reply based on it.

Semantic knowledge structures offer a pattern for linking verbal and nonverbal information. OWL and OWL 2 Web ontology languages [29] are responsible for knowledge representation in terms of creating ontologies for the Semantic Web. OWL 2 design and semantics are strongly connected with Description Logics (DL) [30], while they enable knowledge reusing and inferencing. Axioms, entities, and expressions are some basic notions of OWL ontologies. Understanding the context is a significant issue for exploiting the expressivity of semantic models. Several approaches of applying semantic web technologies have been proposed which superpose interpretation layers in the context. These approaches deal with many issues like pervasive applications [31], natural language interfaces for content disambiguation [32], or fusing contextual information [33].

Only a few studies are associated with connecting ontology-based dialogue entities to domain-specific and social competencies. OntoDM [34] is a system which supports dialogue management based on ontology assistance. In this system, the conversation with the user is driven by keeping past conversations in memory, which is used as a basis to analyse new information. On the other hand, there is also an ontology-based dialogue system which creates domain model representation, applies reasoning rules, suggests the appropriate dialogue representation, and tracks older dialogue states to result in the most appropriate system response [35]. These studies converge in the great contribution that ontologies provide in a dialogue-management process. The justification lies in the fact that all dialogue information which is mapped into the ontologies can be accessed without any limitation in order to select the most appropriate system responses and to apply dialogue strategies.

3. System Architecture

3.1. System Overview

The primary notion that was considered when designing the system's architecture was to exploit medical expertise and cutting-edge monitoring technologies and to combine them into a user-friendly package. Being able to process, collect, and conduct analyses of sensor/camera/user feedback data while simultaneously using the latest advances in ontological frameworks to choose the most fitting course of action, the REA platform interacts seamlessly with patients, providing personalised reactions.

Physical device interfacing and data collection: To achieve unhindered asynchronous communication between servers that perform data collection and the respective IoT devices and to ascertain security in the connections, the platform employs the Message Queuing Telemetry Transport (MQTT) broker of the Amazon Web Services (AWS): Data originating from the IoT devices are handled by a worker process asynchronously, while other system services (service-oriented architecture) are access the data via a Web API. The latter disposes endpoints to configure and send commands to the aforementioned devices via the broker.

Data processing: Due to the versatile nature of the system and the multitude of data sources available at any given time, data aggregation and processing are tackled by dedicated components that operate with specific APIs and generate respective events accordingly. Sensor data originate from physical devices like wearable devices or ambient sensors, which have been customised to fit the project's needs through proprietary firmware implementation. Of particular importance is the sleep sensor placed under patient beds, which assesses sleep quality by monitoring sleep duration, interruptions, person movement, and body posture. The results of the evaluation are stored in the

knowledge base; they update the patient's medical history and could be explored to raise alerts or stimulate the system to proactively communicate with the user, inquiring about possible sleep disturbance causes. Visual data are also gathered using computer vision models aiming to capture events like the activity that the person is currently performing or to determine if the user has fallen. Relevant alerts are raised after data analysis, which evaluates if certain value changes (monitored by the Real Time Operating System (RTOS) firmware implementation) justify such a high-priority action.

Data saving: Data coming from heterogeneous sources (e.g., interface for user profile, communication with the patient, sensor, and camera surveillance) are saved in the knowledge base. All data are converted from a JavaScript Object Notation (JSON) format into a semantic Resource Description Framework (RDF)/OWL format according to the ontologies described in Section 6.1. The dialogue-management component is responsible for providing the flow of information in the semantic component, which is responsible for filling the knowledge base (KB). The moment a user profile is added in the knowledge base, it acts as a pre-trigger to the data-saving procedure, which remains active ever since, thus collecting all information pertinent to the user and mapping it to respective ontological instances.

Self-improvement mechanism: The REA system is automatically improved and updated as long as data are collected and more interactions are conducted between the user and the system. The growing number of data recorded from the sensors increases the system's awareness over the patient's medical condition. The continuous integration of the biometrical indication data and conversations in the knowledge bases helps the inference of information that is not directly fed by any person. In addition, it facilitates and upgrades the decision-making process of the dialogue management by supporting more system actions that are personalized and not completely dependent on the last user utterance.

Machine-human interface: The system manages user interaction with a mix of voice communication and a human-like 3-D model. The voice communication layer converts nonverbal system responses to verbal, assisted by a speech synthesis component, while the 3-D agent supports advanced feedback mechanisms like blinking eyes, lip/speech synchronisation, head movement, or facial expressions to ensure natural communication.

3.2. System Design and Integration

3.2.1. System Conceptualisation

The system's high-level architecture is divided into three different layers (Figure 1):

- The communication understanding layer
- The sensor management layer
- The communication analysis layer

The communication-understanding layer is responsible for receiving any kind of user input and for routing it to the other layers depending on the informational and functionality needs. It also outputs the final system response or action to the user. It comprises the only layer that communicates directly with the user and the central point, where all the information between the user and the system's layers is circulated. It encapsulates modules for transforming any voice input into text and the other way around, and it provides access to data and features that require additional rights to authenticated users (mainly the medical centre staff).

The sensor-management layer collects and distributes data that are recorded by installing sensors in the room or the user's body. Data stored in this layer concern activities and a person's biometrical indications. It is either retrieved by the communication-understanding layer or sent in the content associated with an alert signal triggered in emergencies when a proactive action is necessary. It is also possible to send commands to this layer to perform a move pertinent to the adaptation of equipment such as the patient's bed, where a need to change its angle may arise at any moment. In addition to the sensor systems, the REA platform will also incorporate an advanced computer vision component, responsible for detecting patient falls and activities.

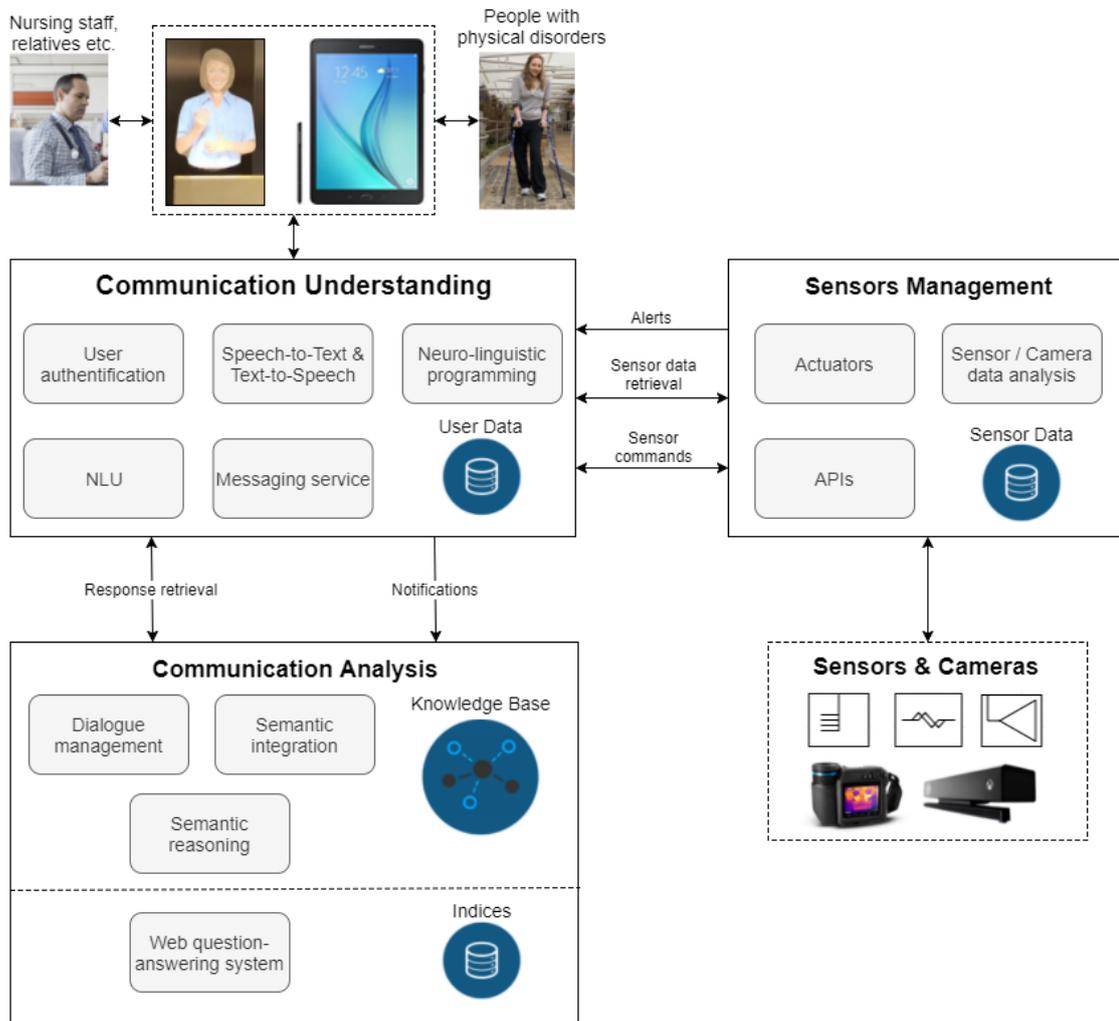


Figure 1. System architecture.

Moreover, a communication analysis layer will be integrated, which will grant the REA agent advanced and complex conversational capabilities. The main module of this layer, the Dialogue Management (DM), semantically represents and interlinks the user utterance against the system’s domain and dialogue models, whereas at the same time, it processes the entire dialogue to generate content that has not been explicitly fed to the platform. It retrieves relevant content for the system action from either the Knowledge Base (KB) or the web-based question answering service. In most cases, the communication-understanding layer obtains the final system response from this layer, regardless of where the accompanying content is retrieved from (e.g., it may come from the sensors layer). Nevertheless, there are some occasions when the DM intervention is not required and the system action is provided directly from the communication-understanding layer. Regardless, at every dialogue turn, the DM is informed of the performed system action, even if it is not generated by this layer, to keep a complete and updated version of the dialogue history.

3.2.2. Sensor Development

The customisation of the sensor systems is made possible by altering suitable commercial sensor chips and devices. These are integrated with a compatible MCU (MicroController Unit) that interface-wise and connectivity-wise (support for wireless connections) can handle the task of measuring the desired values and of sending them to the cloud. The programming of the firmware follows project specifications to ensure the sampling of device and sensor chip values is carried out properly and in regular intervals. To establish that common procedural steps (registering of the value,

proper formatting, and uploading to the cloud) will be followed for every device/sensor, a RTOS will be integrated inside the shared firmware. Due to the fact that it is designed to parallel process numerous tasks, RTOS will also provide a means for speedy completion of the needed procedures.

3.2.3. System Integration

Since the REA platform is still a work in progress, installation in a real environment has not been completed yet. The objective is to integrate developed devices and sensors (see Section 4 for a full list) into both clinical and home settings. The components that are being considered for the first phase include blood pressure sensors, blood sugar sensor, bed management, and camera system. Specifically, gyroscopic sensors on patients' beds will be installed to control the angle of the head and feet part of the beds, which will react to voice commands. Camera functionalities will include fall and crowd detection, activity recognition, and patient temperature readings. This phase will take place in a hospital environment, since it is easier to familiarise users with the system, to monitor them, and get feedback from both caregivers and patients. For the final system integration phase, which will also include home settings, a sleep sensor and a wearable sensor system that will measure temperature, heartbeats, and blood oxygen and that will have the ability to also detect fall will be included.

4. Data Collection

4.1. Monitoring Sensor Types

To ascertain that all data collection and monitoring objectives are met, diverse types of sensors, both ambient and wearable, are being integrated to the platform; commercial sensors have been enhanced with IoT functionalities, while other sensors may be built from scratch to operate according to desired specifications. Therefore, the system receives as input varied types of data, some of which can be directly explored (i.e., temperature measurements) while others require further manipulation to extract meaningful conclusions (i.e., sleep sensor data). Notably, the following sensors and systems will be used:

- Blood pressure sensor: A traditional off-the-shelf blood-pressure system will be enhanced with wireless connectivity, converting it to a smart IoT device.
- Blood sugar sensor: A wireless module will be integrated to a common blood sugar sensor to imbue WiFi capabilities and to enable cloud communication.
- Sleep quality sensor: To assess a user's sleep quality, data from a sensor-integrated sheet that is placed beneath his/her bed is used. The embedded sensors include a Wi-Fi module, which, for the duration of the sleep, uploads data to the REA platform for further analysis. Depending on parameters like body movement, posture, and interruptions, a proactive system intervention could be triggered.
- Vision system: It consists of a depth camera, a thermal camera, and an RGB camera, sending data via a USB connection to a local server, which analyses the relevant footage using computer vision models for crowd detection, fall detection, and activity recognition [36].
- Bed management system: Dual gyroscopic sensors and IoT functionality will be added to patient beds that will allow for remote control of both the head and leg section of each bed using voice commands. Moreover, the same applies to each section's angle, since it is possible for the user, either a patient or a caregiver, to tilt the respective section to the desired position.
- Wearable sensor: Physiological measures like the patient's temperature, heart rate, and blood oxygen levels are monitored by the REA wearable sensor, which takes the form of a bracelet. The integrated accelerometer and gyroscopic sensors give the potential of an early warning in fall emergencies. Through a Bluetooth Low Energy (BLE) gateway, all readings are forwarded to AWS via Wi-Fi.

As already mentioned in Section 3.1, the AWS platform accommodates REA's data collection infrastructure, of which the IoT service is responsible for secure sensor communication via an MQTT

broker; a user-controlled and -parameterised public key infrastructure is available, while AWS Lambda functions tackle sensor MQTT messages before being processed by a task engine, of which the role is to save data in the database or to push notifications to related services. A REST API handles access to the stored database data.

4.2. Information Protection and Privacy

Personal data belonging to the natural persons associated with REA's research, in particular those correlated to patients, escorts, medical, and nursing staff, are protected through appropriate technical and organisational measures. All pertinent actions employed, to the best of our knowledge, were in absolute compliance with the provisions of Regulation (EU) 2016/679 that the council of the 27th of April 2016 and the European Parliament laid out concerning "the protection of individuals with regard to the processing of personal data and on the free movement of such data and the repeal of Directive 95/46/EC" (General Data Protection Regulation) and, finally, with relevant national legislation. More specifically, all required and pertinent security precautions as well as all procedures will be respected and applied to the following stages: Installation and use of the video surveillance system (consisting of thermal, depth, and conventional cameras) and the duration of the storage of data involved. The issue of consent (of all interested parties) is approached in the same perspective to all the above—consent will be received beforehand—with great consideration on the respective personal data, abiding to current legislation, both national and European. The latter encompasses and covers a broad spectrum of areas, such as the commitments of institutions that process personal data, the principles that need to be considered and applied during data processing, the rights attached to implicated parties who provide the relevant data, and finally the directions towards which the relevant rights can be employed.

In particular, personal data security is assured when Secure Sockets Layer (SSL)-based protection methods are preferred and backups are encrypted, as is the case in REA. Moreover, uppermost data handling and infrastructure security is guaranteed (ISO 27001- and PCI DSS-compliant), since access to data centres will be shared between interested parties. Likewise, a Universally Unique Identifier (UUID) accompanies the platform's IoT sensors in a bid to ensure system security. This ID is designated via certificates to these devices using the AWS certificate creation environment to manage secure and encrypted (Transport Layer Security (TLS)) connections to the respective MQTT brokers. As far as data security is concerned, the Sensor Measurement Lists (SenML) IETF standard RFC84286 (<https://tools.ietf.org/html/rfc8428>) is adopted.

5. Language Understanding and Speech Synthesis

Language understanding is a research field which has enjoyed considerable popularity and found application to a broad variety of consumer tasks (e.g., text categorisation, machine translation, and question answering). It focuses on text processing, and in the case of REA, it serves as a mediator between user input and the dialogue manager. Naturally, in order to proceed to the handling of text, user input first needs to be transcribed from audio to text format. This is where speech recognition, an equally popular task, comes into play, being a step of utmost importance; dependant on its performance is the quality of the transcribed text. Despite the exposure that speech recognition has had in recent times, many challenges have not been overcome yet, with issues like background noise, multi-user simultaneous input, homophones, and regional accents proving to be too much of an impediment to be tackled correctly, even by recent platforms. The same applies, though to a lesser extent, to the speech synthesis task, where even small amounts of latency or alterations in voice quality may disgruntle users, since people are sensitive to audio inadequacy.

When examining the prospects of the available platforms that would compose REA's candidate list, the aforementioned issues were considered and the choices were narrowed down to Google, Acapela, and Innoetics for the speech recognition [37] task and to Snips, CMUSphinx, and Nuance for the text-to-speech one. The final criteria were language support, cost, and service quality. It was

imperative that the platform of choice would support both English and Greek natively while also offer competitive subscription rates. Certainly, of paramount importance was the quality of the results that each service offered, since having to process a not well-crafted output which lacks in quality does not do justice to the system no matter how sophisticated and well organised it is. After careful consideration of all heuristics, Google's speech recognition services stood out, largely due to excellent support for Greek, while, concerning the text-to-speech services, Nuance's suite was the most suitable.

Language understanding: In REA, the Google Cloud Speech-to-Text API converts user requests into text form. To process this input and to extract meaningful, system-exploitable information, raw text needs to be processed and augmented by specific NLP techniques that will also disambiguate probable semantic discrepancies. Text processing is handled by Stanford's CoreNLP suite [38], which performs linguistic analysis utilising tools like part-of-speech (POS) parsers, tokenisers, and chunkers to extract dependencies between sentence words, concepts, the underlying relations, named entities, etc. The produced output receives supplementary processing in order to retrieve probable disease/treatment-related relations in user queries by applying a hybrid relation extraction tool [39].

Properly linking the entities and concepts with the appropriate relations is a significant issue for interpreting the semantics of information and confronting complex information requirements. To achieve this goal, a process that connects entities according to their semantics is needed. The process disambiguates the concepts based on their context and links the entities with ontologies and semantic networks such as BabelNet and WordNet. More specifically, REA utilises existing tools for context disambiguation and word sense, e.g., Babelfy (<http://babelfy.org/>), FRED (<http://wit.istc.cnr.it/stlab-tools/fred/>), and DBpedia Spotlight (<https://github.com/dbpedia-spotlight/dbpedia-spotlight>), to link and align information to ontologies. All these tools contribute to transforming the concepts into ontologies, which are able to respond to user requirements concerning semantics and context of information.

Speech synthesis [40]: This is the procedure that vocalises the system's response; language-understanding-generated text, which should be directed to the user in verbal form, is converted by Nuance's vocalizer platform. The model that Nuance employs leverages deep neural network architectures to produce natural-sounding human voice by exploiting great numbers of prerecorded speech data. Human-like levels of expressiveness are achieved by studying not only the currently processed utterance but also its respective context. The same applies for the intonation, which benefits from the exploration of relationships found between text and the respective voice characteristics.

6. Knowledge Representation and Reasoning

6.1. Multimodal Knowledge Representation

The OWL 2 ontology language [29] has been used to combine information from heterogeneous sources and map entities like profiles, activities, medical and verbal communication information, and relationships between them. Based on the class and property assertions in the DL theory [30], observations are mapped and information interconnection is successful. The generated models are complete and offer a variety of benefits, which derive from ontologies like modelling entity relations, connecting and sharing information originating from different sources, and applying reasoning mechanisms.

In the scope of this project, data stem from a plethora of sources such as verbal communication with the user, camera surveillance, sensors, etc. Information has to be represented as domain entities to extract contextual descriptors which satisfy and interpret the context. A basic vocabulary that we use for modelling context types is shown in Figure 2. The model describes events and observations using the `leo:Event` concept of LODÉ [41] and offers core properties to capture temporal entities related to observations.

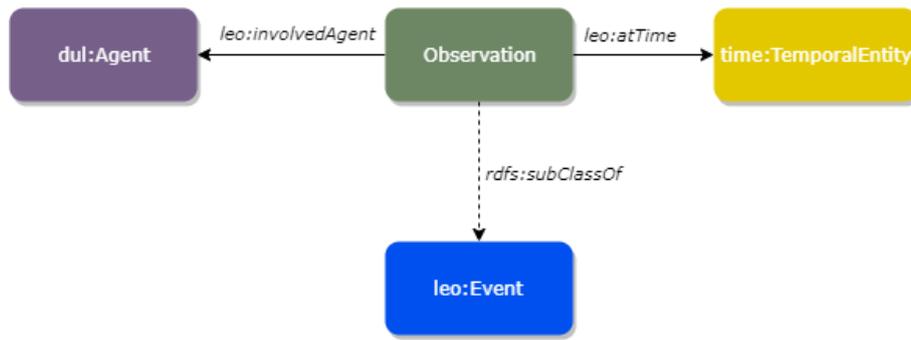


Figure 2. Upper-level domain ontologies for observations.

More complex domain models which consist of information such as user profile, routines, habits, and behavioral aspects are supported. Many existing patterns and ontologies have been reused and extended, like modelling patterns for smart homes [42] and descriptions and situations pattern of the DolceUltralite ontology [43], e.g., for modelling physiotherapy exercises. Figure 3 shows a higher-level representation, offering behavioural aspects as instances of the aspect class and detailed information as instances of the view class.

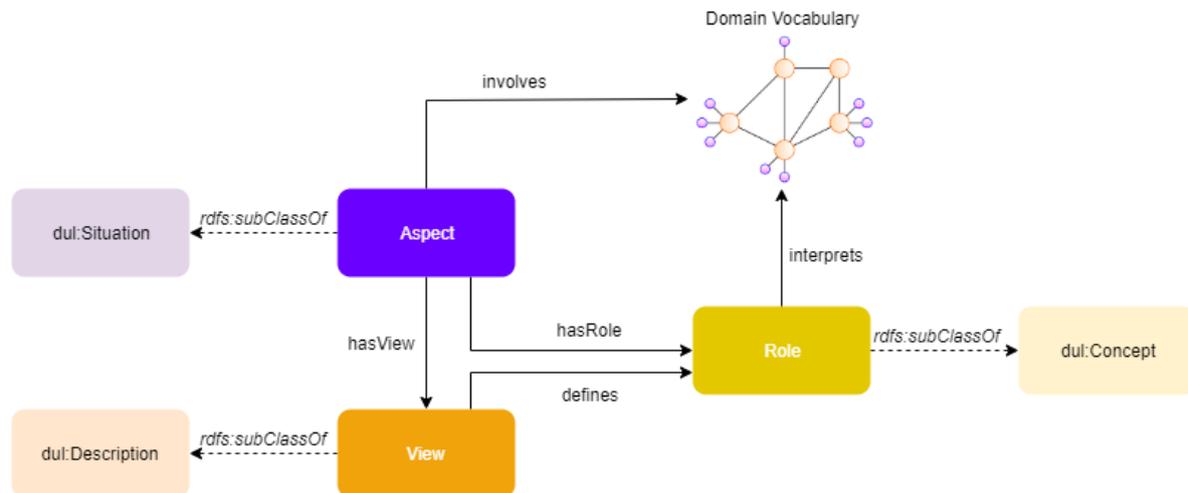


Figure 3. Upper-level pattern for modelling complex situations.

6.2. Data Analytics

The data analytics process is related to the processing of data coming from sensors and includes the following:

- Analysis of wearable sensor data that supports high-level event detection by containing useful information. Surveillance of physical activities is also supported for activities like stress level, quality of sleep metrics, surveillance of lifestyle, and intensity of movement. Each of the activities mentioned above is processed by a different component, which complies with the data formats pertinent to the REA project. For instance, an accelerometer is used in order to detect movement to interpret physical activity and skin conductance, which indicates level of stress. To accomplish this, random forests-based learning techniques are used to apply internal filtering strategies for signal metadata and statistical mechanisms for each measurement [44]. The results return a score relating to the level of stress and intensity of movement in a range from 0 to 5.
- Recognition of human activity is strongly connected with computer vision techniques as mentioned in Reference [36]. The input data, which are coming from depth and IP cameras, include pictures from surrounding areas. This results in the recognition of simple activities like walking, sitting, etc. or more complex ones like eating, washing the dishes, cooking, etc.

6.3. Reasoning

A reasoning mechanism is formed to apply semantic rules and to enrich the available information. Implicit taxonomic relations among concepts are extracted, exploiting OWL 2 and DL reasoning services. A knowledge base is meaningful when satisfaction and consistency checking are applied. In our case, a reasoning mechanism attempts to understand the patient's expressions in order to support the dialogue management with useful information, e.g., about the topic. DL theory helps model the supported topics by matching low-level observations to high-level inferences. Recognising patients' pain is modelled in DL as follows:

$$\text{PainContext} \equiv \text{VerbalContext} \sqcap \exists \text{contains.HurtReference}$$

In this rule, the VerbalContext defines the verbal communication with the patient. The rule defines that, if the communication with the patient contains hurt reference, which is defined by analysing the verbal communication information, then the current context is of the type PainContext.

Recognising Topics

This functionality of the semantic reasoning mechanism is of high importance in the REA project because it contributes to identifying the discussion topic and strongly supports the dialogue-management system. Thus, the component detects the most appropriate response in each user request. It entails a complex procedure, that requires to first contemplate on the available data influx in order to later provide highly customised, situation-aware solutions.

Verbal communication with the user is analysed and clustered into topics using a thematic ontology. This task is led by the reasoner who creates a conversational context and utilises an ontology reasoner to select, in each case, the most appropriate topic from a list of 119 topics. Each topic is defined by more than one case question. A part of the ontology is shown in Figure 4.

The analysis of the verbal communication is strongly connected with the interpretation task. This task leads to the extraction of high-level interpretations using atomic observations. The procedure that is followed consists of creating groups of observations, of selecting the current context, of applying semantic reasoning, and of classifying the context. The temporal extension of observations and the information, which are associated with the domain, result in forming the current context. DiseaseContext topic is defined as follows:

$$\text{DiseaseContext} \equiv \text{Context} \sqcap \exists \text{contains.Disease} \sqcap \exists \text{contains.CareRecipient}$$

The native OWL DL reasoning services are strongly connected with the completeness of data. For instance, even if a small amount of information is missing, i.e., a dependency, the conversational topic is not detected. To overcome such issues, we have to check the quality of data, i.e., the performance of the extracted concepts from verbal communication. Low performance may lead to lack of conversational awareness and topic detection failure. A custom reasoning mechanism has been formed to enhance conversational awareness.

If $T = \{d_1, d_2, \dots, d_n\}$ represents a conversational topic and $i = \{i_1, i_2, \dots, i_l\}$ represents the current context (d_n and i_l are ontology concepts from the domain ontology), the relation below matches a topic T with the current context i , only if they have at least one dependency in common:

$$\text{conv}(T, i) = \frac{|T \cap i|}{|T|}$$

In the aforementioned relation, $T \cap i$ indicates an intersection of a semantic set which returns as a result the mutual components between the two sets. If the two sets do not contain any mutual concepts, then $\text{conv} = 0$, while if there are mutual concepts between the two sets, then $\text{conv} > 0$ and the topic is selected as plausible. The highest conv value determines the most plausible topic.

For instance, if $DiseaseContext = \{Disease, CareRecipient\}$ and $i = \{Disease\}$, then $DiseaseContext$ is a plausible topic having $conv = 0.5$. This functionality supports the topic detection when the input contains missing values, which causes recall increase.



Figure 4. Upper-level classes for the topic hierarchy ontology.

7. Dialogue Management and Web-Based Retrieval

The dialogue-management component is the most significant component in the communication-analysis layer as well as the one that communicates with the other agent architecture layers. It is responsible for the decisions in the majority of the system responses. The main feed of information in this component is taken from the content existing in the described ontologies. Nevertheless, the current DM framework is flexible for integration of different data resources. In REA, apart from the knowledge base, the DM also leverages multimedia content that is situated on the Internet and is provided via the system's question answering service. The DM generates a set of candidate actions which are assessed with the help of a trained decision strategy to conclude on the optimal system response.

7.1. Decision Making

The main functionality of the DM component is to choose the appropriate system response at each dialogue turn. In order to support complex and natural conversations with the user, the component shall have access to data that reside in ontologies. Ontologies render the system capable of performing

advanced functions on the data and understanding the provided information, including user utterances, as humans do.

Furthermore, the DM is capable of handling various interactive events like pauses and hesitations. These ontology-based dialogue-management methods will be part of an extended version of the framework, relevant to how they were implemented in the H2020 KRISTINA [45] research project. The adaptations are related to the requirements that emerged from the REA specifications in order to develop an agent that combines task-oriented and non-task-oriented solutions and is able to address domain-specific (e.g., health care) and social capabilities accordingly. The non-task-oriented actions that are supported by the system and include “chatty”-style utterances produced complementarily to the task-oriented ones to address cases when the intention of the user is not straightforward; they aim to increase the system’s naturalness and to maintain prolonged user engagement in the platform.

In other words, the objective is to develop a system that acts and responds to user requests like a human assistant would do. To this end, it will be able to generate strategies that consider and process (a) the replies retrieved from the knowledge bases and the question answering service, (b) the history of the conversation, (c) any context that can be discovered from the user’s words in the last dialogue turn, and (d) the relevant data connected with the users that also contain their personalised needs. These are the main factors that influence the DM strategy and urges the system to select the most suitable response. When deemed appropriate, system responses may be deduced using the ontology functionalities while being unrelated to the last user utterance. Apart from that, the DM can support interventional or motivational actions, such as alerts, notifications, or reminders. An example that shows the need for these actions is the case of taking care of people with mental diseases; continuous repetition of an identical request in a short time period could indicate some kind of mental crisis, where an alert notification must be sent to the people responsible for the immediate handling of such emergency incidents.

7.2. Web-Based Question Answering Service

This module runs on the background of the dialogue-management module and acts complementary to it as it enhances its knowledge with content existing on the World Wide Web. It is able to respond to questions whenever an answer cannot be retrieved by querying the knowledge base. When the dialogue manager decides that the system response must be addressed by the question answering (QA) service, it sends a request with the user question included and receives as a reply the exact passage from the matching web resource that answers the question.

To setup and support such a service, a preprocessing pipeline must be executed continuously to create and update the resource base to be searched in the question-answering phase. Firstly, websites that address our retrieval needs must be defined. In REA, we decided to work on websites that provide (a) trustworthy health-related information, (b) weather forecasts, (c) newspaper articles, and (d) updates about the most important upcoming events. The preprocessing procedure sequence comprises of the *web crawling and searching* step that discovers webpage addresses connected with the initial websites, the *web scraping* step that extracts the meaningful content out of the aforementioned webpages, and the *indexing* step that prepares the scraped information for quick and efficient retrieval by storing it into specialised data structures (e.g., inverted indices). The result of this pipeline is a resource base that is enriched every time we need to integrate additional knowledge to the agent and is utilised as the searching point of the QA service. This service is an information retrieval application that parses natural language text questions given by the user and finds the top-ranked excerpt in the resource base in terms of relevance to the question.

All the preprocessing steps along with the question-answering service are depicted in Figure 5.

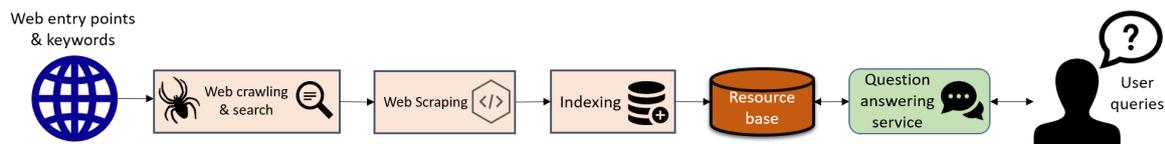


Figure 5. Question answering from web resource framework.

8. Testing Settings and Usage Scenarios

In what follows, the system’s integration settings are introduced and a real use-case example is presented, where each processing step of the system is analysed.

8.1. Testing Environments

System installation has not commenced yet, being an objective pertaining to a subsequent period of the REA project. Nonetheless, testing conditions have been planned beforehand and evaluations will be conducted in both residences and hospitals. The Evexia rehabilitation center (<http://www.evexia.com/en/>) will be the clinical testing ground. Being an experienced service provider, it can accommodate in its 165 beds individuals afflicted by neurological (e.g., Parkinson’s disease) and rheumatic diseases (e.g., rheumatoid arthritis) or postoperative conditions (e.g., intertrochanteric fracture, paraplegia, tetraplegia, etc.).

Evexia will grant ten patient rooms (2 single and 8 double) for the purposes of the REA project. Throughout the duration of the evaluation period, invaluable patient data will be collected by the installed sensors and cameras that will serve both as training material for the platform’s machine learning components and as evaluation metrics, measuring system- interaction quality and satisfaction. The alternate evaluation setting, post-recuperation patient’s residences, will include two individuals who have terminated their stay at the rehabilitation centre and need to continue their treatment at home.

8.2. A Use-Case Scenario

In the following example, we simulate an indicative real-life scenario that involves a probable dialogue between clinical staff and the smart virtual agent concerning a specific patient’s medical condition and clinical history. The platform’s responses are accurate and reliable, facilitating the doctor with immediate access to the desired information via voice commands, leaving his/her hands available to simultaneously perform other tasks, like writing a prescription or physically checking the patient. In Table 1, the relevant dialogue is presented and the consequent subsections illustrate the different components that are activated to support it.

Table 1. Doctor-agent use-case scenario.

Actor	Interaction
(i1) Doctor	REA, could you tell me Mr. Doe’s treatment for tonight?
(i2) REA	Mr. Doe is going to take (Drug 1) and (Drug 2). However, he complains that (Drug 1) has no effect on him.
(i3) Doctor	REA, i would like to know about his medical condition.
(i4) REA	Blood pressure is at normal levels, systolic is 105, and diastolic is 65. His heartbeats are 90 per minute, a little higher than expected. Lastly, his body temperature is 36.6 degrees Celsius.
(i5) Doctor	That’s all REA. Thank you.
(i6) REA	Enjoy the rest of the day Dr. Shepherd!

8.2.1. Language Understanding

The doctor's request is transformed from voice input to text transcription and is eventually broken down to valuable keywords and dependencies. These include concepts, named entities, and the relations that bind them. A good example is sentence (i1), where the platform extracts the concept "treatment" and the named entity "Doe". Babelnet is then employed to retrieve the respective synset (treatment -> <https://babelnet.org/synset?word=bn:00047235n>), while the patient's profile is accessed using the extracted named entity.

8.2.2. Reasoning

After the conceptual analysis of the text of the request, the semantic meaning behind the terms needs to be detected to support the dialogue-management process. More specifically, the reasoning mechanism supports the DM in the topic selection issue. In the following example, we define an InformMedicalCondition request when the verbal communication context contains both medication and state contexts.

$$\text{InformMedicalCondition} \equiv \text{VerbalContext} \sqcap \exists \text{contains.Medication} \sqcap \exists \text{contains.Condition}$$

8.2.3. Dialogue Management

The first step of DM is to identify the discussion topic (AskForPatientMedication) as well as the main entities; the named entity "Doe" and the corresponding "Day" and "Time" instances that arise from the "tonight" keyword. Then, it consults the appropriate data sources and services for determining the optimal system response. In the first dialogue turn, DM needs data from the KB that are related to "Mr. Doe"'s treatment at a specific time. Thus, an indicative KB reply could be formed as follows:

```
{
  "type": "MedicationAdministration",
  "": [
    {
      "drug": "Drug 1",
      "schedule": "Night"
    },
    {
      "drug": "Drug 2",
      "schedule": "Night"
    }, ...
  ]
}
```

As aforementioned in the DM description, the module can produce candidate actions that are irrelevant to the current conversational status and rely on the dialogue history. If the patient has already shown his/her dissatisfaction about the drugs many times (the exact number must be specified using domain expert knowledge), the system shall notify the doctor as well. Therefore, in the current example, the predefined topic-based response will be sent to the doctor along with a notification showing the patient's complaints. The selected system responses along with their relevant content (which has the form of concepts) are given to the communication-understanding module to generate the phrases that the user will receive as reply (i2).

The execution steps of the DM module are similar in all the dialogue turns. A noteworthy fact is that the module is capable of recognising the context of pronouns like "his" in the i3 turn, where it is properly associated with "Mr. Doe", which in such cases is an essential capability for detecting and creating the correct system response.

8.3. Preliminary User-Centred Evaluation

The objective of the preliminary user-centred evaluation is to obtain early feedback from end-users of the system to fuel the development before the actual deployment in real environments. While the system currently is not considered fully operational, a compromised pre-alpha version with specific functionalities is available for testing and evaluation in controlled laboratory settings. The operational modules only provide limited options, which restricts significantly the potential outcome; however, the provided feedback is enough to influence the platform's direction.

As described also in Section 3.2, the modules have been integrated in the REA framework; however, they are not yet part of the pilots. Instead, internal IT and end-users were invited to test the current implementation through guided conversations with the system in accordance with Good Clinical Practice (GCP). The process took about 15 minutes per participant (6 participants in total) that filled in a five-point scale questionnaire (1—completely agree, 5—completely disagree) with assistance of personnel. The questionnaire was compiled and underwent two review processes within the consortium. Sample questions are depicted in Table 2.

Table 2. Example questionnaire responses.

Question	Mean (SD)
Provides too little information on treatments	2.31 ± 0.85
Provides too little information on problems	2.35 ± 0.89
Addresses user's information requests	3.21 ± 1.21
Contradictory information is returned	3.09 ± 1.56
It takes too long to respond	2.11 ± 0.9
There are too many "can you please repeat the question?" responses	1.73 ± 0.95

Due to the limited number of participants, the results may not be considered as representative and must be interpreted carefully. However, the questionnaires revealed the following critical aspects:

- The response of the system is too often "Can you repeat the question?" The topic detection task depends solely on the results coming from speech and language analysis, which are used by the underlying reasoner to classify utterance contexts in the topic hierarchy. The current implementation is not able to handle missing information and uncertainty. Therefore, the absence of a term from the input hampers the detection of the correct topic.
- It takes too long to provide a response. The average response time of the framework was 2.4 s that, based on users' feedback, needs to be further improved. Although ontological reasoning imposes some inherent scalability issues, we plan to investigate more scalable reasoning schemes, such as to reduce, if possible, the expressivity of the models.

On the positive side, users highlighted the fact that the framework addresses in principle the information requests of the users and that no contradictions are returned, taking into account the profile information that the system has been initialized with.

9. Conclusions

The specific project, which is still in progress, outlined in the current paper, entails the platform's presently ongoing development covering all main components of the REA system. Every single module responds to a variety of differentiated technological demands that formulate a whole system of which the primary goal is to promote and assist the natural communication between a machine and a human pertinent to patients who are currently under rehabilitation or trying to recuperate. Caregivers and the final consumer both in home and hospital settings will have the chance to assess the project's outcomes through expanded tests performed during the trial of the prototype.

Additionally, when it comes to constant system utilisation and engagement, user acknowledgement of its utility plays a decisive role. Perceptive and easily accessible means that establishing machine and human interaction has been proven to reinforce the concept of self-management, to inspire self-confidence, and to promote the benefits of a more relaxed way of living, especially when it comes to the residential context. Therefore, our primary goal is to concoct an elaborate plan to oversee and evaluate all different and gradual stages of user's involvement, such as approval, easy access, and handling of the demonstrated technologies. We strongly believe that this method will help us to distinguish and determine risk factors involved in user acceptance/approval. Consequently, the above will provide a strong case to various stakeholders who intend to use and rely upon the specific service. Also, this study involves the design, demonstration, and implementation of both key indicators and guidelines that will measure and evaluate the platform effects upon patients. Finally, there is strong intention to continuously communicate REA's objectives to clinicians and employers in order to motivate their employees to actively engage through all project stages; to not be restricted solely towards the scope of capturing, specifying, and designing requirements; to engage during module integration and verification of results.

Author Contributions: Conceptualization, T.M. and G.M.; methodology, T.M. and G.M.; software, S.S.; validation, S.S., E.K., and D.T.; formal analysis, L.P.; investigation, G.A.; resources, C.E.; data curation, G.A.; writing—original draft preparation, T.M., G.M., S.S., C.E., M.R., D.T., and E.K.; writing—review and editing, T.M. and G.M.; visualization, S.S.; supervision, G.M.; project administration, S.V.; funding acquisition, I.K.

Funding: This research received no external funding.

Acknowledgments: This research has been cofinanced by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship & Innovation, under the call RESEARCH-CREATE-INNOVATE (project code: T1EDK-00686).

Conflicts of Interest: The authors declare no conflict of interest. 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.

References

1. Bates, D.W. The quality case for information technology in healthcare. *BMC Med. Inform. Decis. Mak.* **2002**, *2*, 7. [[CrossRef](#)] [[PubMed](#)]
2. Wickramasinghe, N.; Fadlalla, A.; Geisler, E.; Schaffer, J. A framework for assessing e-health preparedness. In Proceedings of the AMCIS 2004 Proceedings, New York, NY, USA, 6–8 August 2004.
3. Marohn, D. Biometrics in healthcare. *Biom. Technol. Today* **2006**, *14*, 9–11. [[CrossRef](#)]
4. Bellegarda, J.R.; Monz, C. State of the art in statistical methods for language and speech processing. *Comput. Speech Lang.* **2016**, *35*, 163–184. [[CrossRef](#)]
5. Ly, K.H.; Ly, A.M.; Andersson, G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interv.* **2017**, *10*, 39–46. [[CrossRef](#)]
6. Tanaka, H.; Negoro, H.; Iwasaka, H.; Nakamura, S. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLoS ONE* **2017**, *12*, e0182151. [[CrossRef](#)]
7. Savino, J.A.; Latifi, R. Hospital and Healthcare Transformation over Last Few Decades. In *The Modern Hospital*; Springer: Berlin, Germany, 2019; pp. 23–29.
8. Ravindranath, P.A.; Hong, P.; Rafii, M.S.; Aisen, P.S.; Jimenez-Maggiore, G. A step forward in integrating healthcare and voice-enabled technology: Concept demonstration with deployment of automatic medical coding model as an amazon “alexa” skill. *Alzheimer's Dement. J. Alzheimer's Assoc.* **2018**, *14*, P955. [[CrossRef](#)]
9. Yaghoubzadeh, R.; Kramer, M.; Pitsch, K.; Kopp, S. Virtual agents as daily assistants for elderly or cognitively impaired people. In *International Workshop on Intelligent Virtual Agents*; Springer: Berlin, Germany, 2013; pp. 79–91.
10. Fadhil, A. Beyond patient monitoring: Conversational agents role in telemedicine & healthcare support for home-living elderly individuals. *arXiv* **2018**, arXiv:1803.06000.

11. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.; et al. Conversational agents in healthcare: A systematic review. *J. Am. Med. Informat. Assoc.* **2018**, *25*, 1248–1258. [[CrossRef](#)]
12. Bickmore, T.W.; Trinh, H.; Olafsson, S.; O’Leary, T.K.; Asadi, R.; Rickles, N.M.; Cruz, R. Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant. *J. Med. Internet Res.* **2018**, *20*, e11510. [[CrossRef](#)]
13. Siddharthan, A. Ehud Reiter and Robert Dale. Building Natural Language Generation Systems. Cambridge University Press, 2000. \$64.95/£ 37.50 (Hardback). *Nat. Lang. Eng.* **2001**, *7*, 271–274. [[CrossRef](#)]
14. Gatt, A.; Krahmer, E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *J. Artif. Intell. Res.* **2018**, *61*, 65–170. [[CrossRef](#)]
15. Perera, R.; Nand, P. Recent advances in natural language generation: A survey and classification of the empirical literature. *Comput. Inform.* **2017**, *36*, 1–32. [[CrossRef](#)]
16. Ze, H.; Senior, A.; Schuster, M. Statistical parametric speech synthesis using deep neural networks. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Vancouver, Australia, 26–31 May 2013; pp. 7962–7966.
17. Tokuda, K.; Nankaku, Y.; Toda, T.; Zen, H.; Yamagishi, J.; Oura, K. Speech synthesis based on hidden Markov models. *Proc. IEEE* **2013**, *101*, 1234–1252. [[CrossRef](#)]
18. Hough, J. Incremental semantics driven natural language generation with self-repairing capability. In Proceedings of the Second Student Research Workshop associated with RANLP, Hissar, Bulgaria, 13 September 2011; pp. 79–84.
19. Oord, A.v.d.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. Wavenet: A generative model for raw audio. *arXiv* **2016**, arXiv:1609.03499.
20. Shen, J.; Pang, R.; Weiss, R.J.; Schuster, M.; Jaitly, N.; Yang, Z.; Chen, Z.; Zhang, Y.; Wang, Y.; Skerrv-Ryan, R.; et al. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgart, AL, Canada, 15–20 April 2018; pp. 4779–4783.
21. Zen, H.; Sak, H. Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 4470–4474.
22. Pollet, V.; Zovato, E.; Irhimeh, S.; Batzu, P.D. Unit Selection with Hierarchical Cascaded Long Short Term Memory Bidirectional Recurrent Neural Nets. In Proceedings of the INTERSPEECH, Stockholm, Sweden, 20–24 August 2017; pp. 3966–3970.
23. Zigel, Y.; Litvak, D.; Gannot, I. A method for automatic fall detection of elderly people using floor vibrations and sound—Proof of concept on human mimicking doll falls. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 2858–2867. [[CrossRef](#)] [[PubMed](#)]
24. Mirmahboub, B.; Samavi, S.; Karimi, N.; Shirani, S. Automatic monocular system for human fall detection based on variations in silhouette area. *IEEE Trans. Biomed. Eng.* **2012**, *60*, 427–436. [[CrossRef](#)]
25. Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J. Monocular 3D head tracking to detect falls of elderly people. In Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society, New York, NY, USA, 30 August–3 September 2006; pp. 6384–6387.
26. Pilcher, J.J.; Ginter, D.R.; Sadowsky, B. Sleep quality versus sleep quantity: Relationships between sleep and measures of health, well-being and sleepiness in college students. *J. Psychosom. Res.* **1997**, *42*, 583–596. [[CrossRef](#)]
27. Heit, J.; Srinivasan, S.; Benitez, D.; Andrews, B.W. Device and Method to Monitor, Assess and Improve Quality of Sleep. U.S. Patent 8,348,840, 8 January 2013.
28. Lison, P. Structured Probabilistic Modelling for Dialogue Management. Ph.D. Thesis, Department of Informatics Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway, October 2014.
29. Graua, B.C.; Horrocks, I.; Motika, B.; Parsiab, B.; Patel-Schneider, P.; Sattler, U. Web Semantics: Science, Services and Agents on the World Wide Web. *Web Semant. Sci. Serv. Agents World Wide Web* **2008**, *6*, 309–322.
30. Baader, F.; Calvanese, D.; McGuinness, D.; Patel-Schneider, P.; Nardi, D. *The Description Logic Handbook: Theory, Implementation and Applications*; Cambridge University Press: Cambridge, UK, 2003.

31. Ye, J.; Dasiopoulou, S.; Stevenson, G.; Meditskos, G.; Kontopoulos, E.; Kompatsiaris, I.; Dobson, S. Semantic web technologies in pervasive computing: A survey and research roadmap. *Pervasive Mob. Comput.* **2015**, *23*, 1–25. [CrossRef]
32. Damljanović, D.; Agatonović, M.; Cunningham, H.; Bontcheva, K. Improving habitability of natural language interfaces for querying ontologies with feedback and clarification dialogues. *Web Semant. Sci. Serv. Agents World Wide Web* **2013**, *19*, 1–21. [CrossRef]
33. Dourlens, S.; Ramdane-Cherif, A.; Monacelli, E. Multi levels semantic architecture for multimodal interaction. *Appl. Intell.* **2013**, *38*, 586–599. [CrossRef]
34. Altinok, D. An ontology-based dialogue management system for banking and finance dialogue systems. *arXiv* **2018**, arXiv:1804.04838.
35. Wessel, M.; Acharya, G.; Carpenter, J.; Yin, M. OntoVPA—An Ontology-Based Dialogue Management System for Virtual Personal Assistants. In *Advanced Social Interaction with Agents*; Springer: Berlin, Germany, 2019; pp. 219–233.
36. Avgerinakis, K.; Briassouli, A.; Kompatsiaris, I. Recognition of activities of daily living for smart home environments. In Proceedings of the 9th International Conference on Intelligent Environments, Athens, Greece, 18–19 July 2013; pp. 173–180.
37. Toshiwal, S.; Kannan, A.; Chiu, C.C.; Wu, Y.; Sainath, T.N.; Livescu, K. A comparison of techniques for language model integration in encoder-decoder speech recognition. In Proceedings of the IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 18–21 December 2018; pp. 369–375.
38. Manning, C.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.; McClosky, D. The Stanford CoreNLP natural language processing toolkit. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MA, USA, 23–25 June 2014; pp. 55–60.
39. Mavropoulos, T.; Liparas, D.; Symeonidis, S.; Vrochidis, S.; Kompatsiaris, I. A Hybrid approach for biomedical relation extraction using finite state automata and random forest-weighted fusion. In *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing, Budapest, Hungary, 17–23 April 2017*; Springer: Berlin, Germany, 2017; pp. 450–462.
40. Jia, Y.; Zhang, Y.; Weiss, R.; Wang, Q.; Shen, J.; Ren, F.; Nguyen, P.; Pang, R.; Moreno, I.L.; Wu, Y.; et al. Transfer learning from speaker verification to multispeaker text-to-speech synthesis. In *Advances in Neural Information Processing Systems*; MIT PRESS: Cambridge, MA, USA, 2018; pp. 4480–4490.
41. Shaw, R.; Troncy, R.; Hardman, L. Lode: Linking open descriptions of events. In Proceedings of the Asian Semantic Web Conference, Shanghai, China, 7–9 December 2009; Springer: Berlin, Germany 2009; pp. 153–167.
42. Meditskos, G.; Dasiopoulou, S.; Kompatsiaris, I. MetaQ: A knowledge-driven framework for context-aware activity recognition combining SPARQL and OWL 2 activity patterns. *Pervasive Mob. Comput.* **2016**, *25*, 104–124. [CrossRef]
43. Gangemi, A.; Mika, P. Understanding the semantic web through descriptions and situations. In Proceedings of the OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”, Catania, Italy, 3–7 November 2003; Springer: Berlin, Germany, 2003; pp. 689–706.
44. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
45. Pragst, L.; Miehle, J.; Minker, W.; Ultes, S. Challenges for adaptive dialogue management in the KRISTINA project. In Proceedings of the 1st ACM SIGCHI International Workshop on Investigating Social Interactions with Artificial Agents, Glasgow, UK, 13–17 November 2017; pp. 11–14.

