symmetry

Article

Multilingual Conversational Systems to Drive the Collection of Patient-Reported Outcomes and Integration into Clinical Workflows

Izidor Mlakar [1,*], Valentino Šafran [1], Daniel Hari [1], Matej Rojc [1], Gazihan Alankuş [2], Rafael Pérez Luna [3] and Umut Ariöz [1,*]

[1] Faculty of Electrical Engineering and Computer Science, University of Maribor, 2000 Maribor, Slovenia; valentino.safran@um.si (V.Š.); daniel.hari@um.si (D.H.); matej.rojc@um.si (M.R.)
[2] Emoda Software, 35430 İzmir, Turkey; gazihan@emodayazilim.com
[3] Dedalus Company, 28212 Madrid, Spain; rafael.perez.luna@dedalus.group
[*] Correspondence: izidor.mlakar@um.si (I.M.); umut.arioz@um.si (U.A.)

Citation: Mlakar, I.; Šafran, V.; Hari, D.; Rojc, M.; Alankuş, G.; Pérez Luna, R.; Ariöz, U. Multilingual Conversational Systems to Drive the Collection of Patient-Reported Outcomes and Integration into Clinical Workflows. *Symmetry* **2021**, *13*, 1187. https://doi.org/10.3390/sym13071187

Academic Editors: Juan Luis García Guirao and Basil Papadopoulos

Received: 28 May 2021
Accepted: 25 June 2021
Published: 1 July 2021

Abstract: Patient-reported outcomes (PROs) and their use in the clinical workflow can improve cancer survivors' outcomes and quality of life. However, there are several challenges regarding efficient collection of the patient-reported outcomes and their integration into the clinical workflow. Patient adherence and interoperability are recognized as main barriers. This work implements a cancer-related study which interconnects artificial intelligence (spoken language algorithms, conversational intelligence) and natural sciences (embodied conversational agents) to create an omni-comprehensive system enabling symmetric computer-mediated interaction. Its goal is to collect patient information and integrate it into clinical routine as digital patient resources (the Fast Healthcare Interoperability Resources). To further increase convenience and simplicity of the data collection, a multimodal sensing network is delivered. In this paper, we introduce the main components of the system, including the mHealth application, the Open Health Connect platform, and algorithms to deliver speech enabled 3D embodied conversational agent to interact with the cancer survivors in five different languages. The system integrates cancer patients' reported information as patient gathered health data into their digital clinical record. The value and impact of the integration will be further evaluated in the clinical study.

Keywords: patient reported outcomes; multimodal sensing; embodied conversational agents; artificial intelligence; spoken language interfaces; cancer survivors; FHIR

1. Introduction

Patient-reported outcomes (PROs) are a type of patient-gathered health data (PGHD), collected from patients to help address a health concern [1]. Since they represent self-reports from every-day life, PROs are an interesting data source in healthcare due to their usage for improving patients' experience, quality of life, and participation of the patient in the clinical workflow [2]. With the technological advance, PROs have become a complementary data source to telemonitoring [3], data mining, imaging-based AI techniques [4–7] as PROs are more sensitive to treatment-related differences and give patients a voice [8]. The knowledge domains of clinical specialties are expanding rapidly and due to the sheer volume and complexity of data, clinicians often fail to exploit it [9].

The first use of patient-reported outcomes was proposed in 1988 [10]. The study highlighted the different concepts on how to collect patient data, compared different areas of possible use of PROs, and defined the directions for further development of PROs [11]. Given the overall technological advances of the era, patient outcomes were collected mostly face-to-face, using paper-written forms. Those forms were then added to paper-form health records (HRs). With the significant advance of information and communication technologies (ICT), i.e., the internet, and mobile technologies, the HRs are slowly but steadily being digitalized. Similarly, the electronic collection of PROs is on uptake [12]. Recent studies confirm the acceptability and efficiency of electronic questionnaire apps on smartphones [13,14]. Electronic PROs, supported by machine learning and other branches of artificial intelligence, can significantly improve drop-out- and acceptance-rates, as well as patient and clinical 'satisfaction' [15–17]. A clear example of how PROs and patient-gathered health data (PGHD) can improve quality of life (QoL) is the domain of ambient assisted living (AAL). In general, AAL environments exploit smart home products, mobile devices, smart watches, and software applications to sense data from the rich source of PGHD in the individual's every-day environment [17,18]. Advances in speech and natural language processing (NLP) technologies has opened up a new paradigm in more personalized and human-like interaction, i.e., symmetric multimodality. With spoken language interfaces, chatbots, and enablers, the conversational intelligence became an emerging field of research in man-machine interfaces. The conversational agents have the potential to play a significant role in healthcare; from assistants during clinical consultations, to supporting positive behavior changes, or as assistants in living environments helping with daily tasks and activities [19,20].

The advances in interactive techniques may have significant impact on patient-adherence and even long-term sustainable quality of results over time, however, the lack of standardization, interoperability, and integration of PGHD have been recognized as the main challenges which are slowing down the uptake of PGHD in the clinical workflows [21,22]. Especially integration of PGHD data in clinical decision making and workflows still poses a problem in healthcare. Unified representation of electronic health-records (EHRs) remains the main issue in the interoperability of electronic health records in general. Fast Healthcare Interoperability Resources (FHIR) is the promising solution for healthcare integration and interoperability. FHIR enables the creation, editing, deletion, and exchange of definitions of medical sources for specific profiles such as Patient, Observation, Questionnaire, and more than 140 other types of resources [23]. HAPI FHIR, an open-source implementation of the FHIR in Java, allows converting between FHIR sources and other application data. Moreover, it allows access to external server resources, and also enables external applications to access or edit the application data.

To sum up, to get the highest contribution from PGHD and PROs, the following points must be considered: (i) 'how to efficiently collect data from patients?', (ii) 'the cost and time for collecting PROs?', (iii) 'how to integrate data into clinical workflow?', and (iv) 'how to enable proper interpretation by the clinicians?'. This study has been performed within a Horizon 2020 project (PERSIST, https://projectpersist.com/, last accessed 19 June 2021). In this study, we propose a holistic system for collecting PROs remotely via both chatbots and ECAs. Further we propose to integrate PROs into the clinical workflow by using FHIR. The FHIR server is located at the open health connect (OHC) platform. All traffic is managed by a multimodal sensing network (MSN) composed of different microservices, such as Text-to-speech (TTS), ECA and automatic speech recognition (ASR). To address points (i) and (ii) we offer a fully symmetric model of interaction supporting speech, gesture, facial expression on both input and output. To address points (iii) and (iv), the FHIR methodology is delivered as an enabler for efficient integration and a fully functional FHIR server to aggregate the PGHD data along with other EHRs, which are integrated into the clinical workflow.

This paper is structured as follows: In Section 2, related works and the differences of our study will be explained. The general platform is described in Section 3. The methodology and the results are described in Sections 4 and 5, respectively. In Section 6, the main contributions are discussed, and the paper ends with the conclusions.

2. Related Works

The paradigm of value-based healthcare represents an important shift towards more efficient and more effective medical care [24]. It creates and environment where teambased care, and meaningful physician-patient patterns are put into center. However, it requires new sources of data to improve shared decision making and enable personalized decision-making. Conversational intelligence is one of the main digital technologies that can significantly contribute to patient activation and engagement. Overall, the technology is based on spoken language technologies (SLT), i.e., NLP, ASR and TTS, that enables machines to interact with humans over mobile or web platforms [25]. In healthcare, the first adaptation of these technologies was proposed in early 1966, with ELIZA [26]. Since then, the NLP and SLT has progressed significantly. The conversational agents have been used to solve complex tasks, such as booking tickets, fetching the result from API and, therefore, acting as customer service agents [27]. In the context of healthcare, conversational agents are intended to provide patients with personalized health and therapy information, relevant products and services, to connect them with health care providers as well as suggest diagnoses and recommended treatments based on patient symptoms and reports. Having properties such as cost-effectiveness, multilingual communication, and 24/7 availability, make conversational agents useful for patients with medical concerns outside of their doctor's operating hours. Studies also reported that patients perceive conversational agents as safer interaction partners than human physicians and are consequently willing to disclose more medical information and report more symptoms to them [28].

In the medical domain, particularly in oncology setting, conversational intelligence focuses primarily on (speech-enabled) chatbots [29] to contribute to screening (i.e., iDecide [30]), improving mental health state through managing psychological distress [31–33] and lifestyle changes [34]. Overall, chatbots have been proven as an enabler for active patient engagement, adherence and satisfaction increase [35,36]. "The self-reporting aspect delivered with the mobile application provides benefits that might otherwise be difficult to obtain" [36]:6689. However, the chatbots have yet to tackle the long-term adherence with sustainable quality of the reported data [37]. In [36], active use of the technology drops after 14 days. Patients' understanding (i.e., familiarity), their ability to remember the details and perceived trustworthiness represent the main factors of patient adherence [38].

Instead of delivering merely a chatbot, in the proposed system, an ECA is presented. ECAs can further increase the long-term adherence by engaging with users in more diverse interaction significantly enriched by incorporating non-verbal communication [37]. One of the earliest definitions of ECA is "more or less autonomous and intelligent software entities with an embodiment used to communicate with the user" [39]. ECAs can deliver a system with symmetric multimodality with speech, gesture, facial expression on both the input and the output side. The main three components of ECAs are user interfaces for communication with the ECA; computer modelling structure to make the ECA react emphatically; and embodiment (visual representation) for communication with users. Embodiments can be designed as virtual human characters [40], animals [41], or robots [42]. In general, the fully symmetric interaction opens up the opportunity to introduce human-like qualities, significantly improving the believability of the interfaces [43]. The main areas of ECAs in healthcare are the treatment of mood disorders, anxiety, psychotic disorders, autism, and substance use disorders [44]. In a review of Kramer [17] about the design and evaluation of the ECAs for healthcare, ECAs proved a promising tool for persuasive communication in healthcare. And, in another review study [42], technological and clinical possibilities of less complex ECAs were investigated and shown as a solution for routine applications in the means of rapid development, testing, and application. The design features of ECAs for healthcare were investigated by Stal [45] who found that the agents' speech and/or textual output, as well as its facial and gaze expressions are the most used features for ECAs. Previous ECA studies for healthcare mainly focused on physical activity [46–48], nutrition [49,50], stress [34], blood glucose monitoring [41], and sun protection [51]. Overall, most of related research on ECAs, however, focuses on speech, facial, and gaze expressions as the main design features [45]. Most of the ECAs in healthcare are 2D based. Their gestures (as part of the non-verbal communication channel) and appearance are most often not considered as main design features. In fact, Kramer [17] reviewed 20 ECA studies to compare their functionalities and appearances. For the appearance of ECA, most of the studies used similar virtual human characters like middle-aged African American women. Moreover, only 3 studies actually addressed gestures. As a contrast we offer two fully ECAs, a male and a female, capable to interact with patients in six languages: English, French, Latvian, Slovenian, Spanish, and Russian. Both can not only represent facial expressions but exploit gestures to enhance user experience by regulating communicative relationships, support verbal counterparts, and maintain certain degree of clarity in the discourse.

Although highly sophisticated, the (embodied) conversational agents are designed as a prototype (proof-of-concept) and the actual contribution to health-related outcomes is evaluated without relevant statistical significance [52]. Further, the interoperability and integration of data collected in the clinical workflow is not considered. However, meaningful use of collected data will significantly contribute to patient adherence. Sayeed et al. [53] describe an approach to create a patient-centered health system that is based on the FHIR standard and patients/clinicians' applications that can make requests and reports of HL7 FHIR resources. Following the same baseline workflow, i.e., the collection of PGHD and forming of FHIR resources, the proposed systems aggregates combine FHIR resources with MSN to offer a fully connected and integrated approach of collecting and integrating PGHD in clinical workflow.

To sum up, the main contributions of proposed system are:

- A multilingual, fully articulated ECA implementing symmetric interaction in 6 languages and with male and female representation
- A micro-service-based sensing network to collect patient information further supported by patient and clinician mHealth application
- A holistic approach towards interoperable and fully integrated PGHD

3. PERSIST Platform for Efficient Collection of PROs

3.1. The PERSIST Sensing Network

The main building blocks in MSN are outlined in Figure 1. The MSN consist of Apache Camel that implements the REST API, ActiveMQ Artemis, and Apache Kafka to implement efficient machine-machine communication among the various service blocks. ActiveMQ Artemis implements the MQTT Broker, whereas Apache Kafka is exploited to deliver an efficient microservice architecture. Apache Camel can act as a router by having the ability to convert synchronous messages to asynchronous ones and vice versa. Apache Camel can also run as a Spring Boot application that provides REST API endpoints for HTTP requests. The MQTT broker is an intermediate between OHC and mHealth App. In this case, the mHealth App is delivered as an MQTT client subscribed to ActiveMQ Artemis. Microservices are interconnected using Kafka topics and HTTP APIs. For Kafka services, asynchronous communication is used where predefined topics for each language are supported and synchronous communication is used for Rasa chatbot that uses HTTP REST requests over Camel REST endpoints API.

Considering the connections, with the mHealth app Figure 2 outlines two types of connections. The first one is the synchronous connection, communication over the secured application protocol HTTPS REST used for questionnaires requests and responses. The second one is the asynchronous connection with the MQTT protocol that uses MQTT topics. Connections to the OHC platform are also using synchronous HTTPS REST protocol. For MSN internal connections, besides REST and MQTT, Camel Java Messaging Service (JMS) and Kafka topics are used.

3.2. mHealth Application

In the study, patients and clinicians have separate mHealth applications (Figure 3). The patient mHealth application is mainly used for data gathering and trends monitoring; the clinician mHealth application is mainly used for patient monitoring and specifying the patient's care plans. Both mobile applications were developed by the company Emoda.

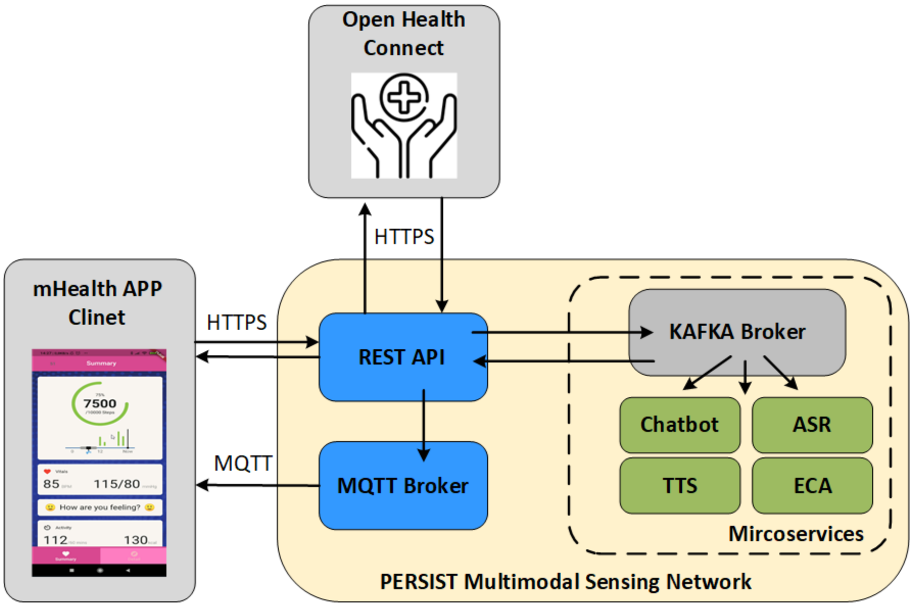


Figure 1. The architecture of the PERSIST MSN.

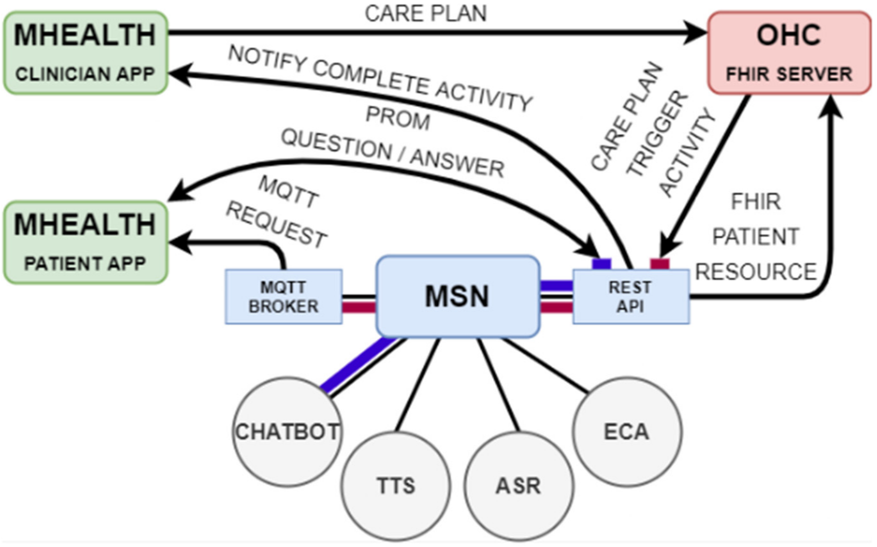
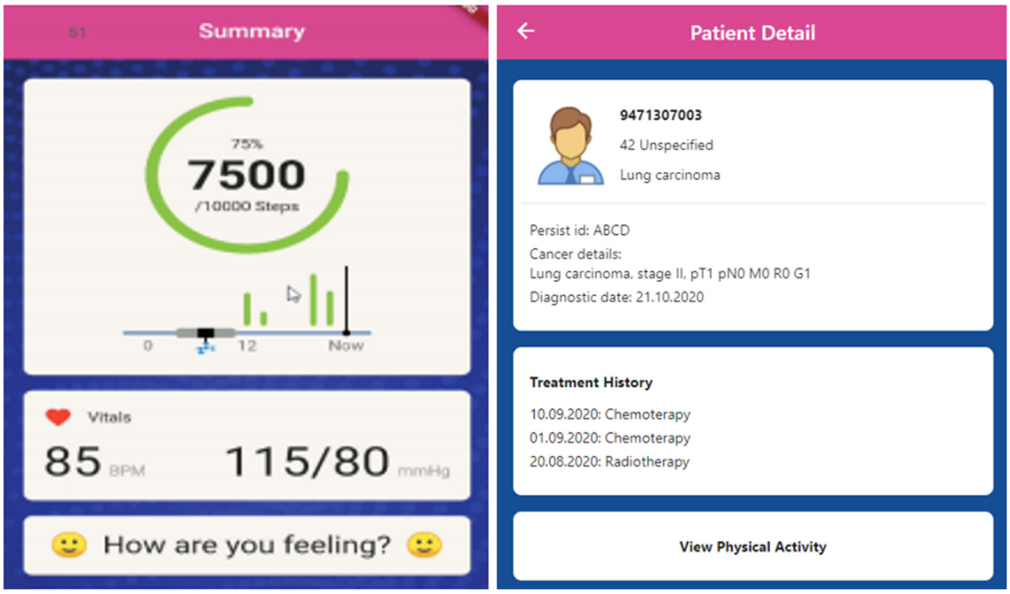


Figure 2. Machine-to-Machine communication in PERSIST MSN.

The patient application has functionalities such as mood selection, diary recordings, reading of selected articles by clinicians, answering the questionnaires, receiving messages from the clinician, and clinician appointment scheduling as the main functionalities of the application. For the clinician application, clinicians have options to see all the patients' lists and their clinical details. They are able to create a new patient record and edit or delete an already existing one. Also, clinicians can create appointments, see the calendar, receive alerts for specific patients, and send/receive messages from patients. The mHealth App can use both synchronous and asynchronous protocols. While the synchronous REST protocol is used for communication with OHC and MSN REST OpenAPI (Swagger) endpoints, asynchronous MQTT protocol is used for receiving notifications.

3.3. OHC FHIR Server

The OHC platform is the complete integration and streaming platform for large-scale distributed environments provided by Dedalus. OHC is a digital health platform that helps unlock isolated data. By consolidating data in a standardised (FHIR) format from across a broad range of systems of record, OHC enables innovation through near real time access to longitudinal patient records. Our deliberately open and modular architecture allows OHC to be adaptive to the specific business outcomes. OHC enables all the interfaces to be connected to and make decisions across disparate data sources in real-time. The OHC Digital Health Platform comprises a set of components, as depicted in the conceptual/logical architectures. The OHC solution is flexible and can be deployed on-premise (private data center) or via cloud in environments like Azure or AWS. OHC provides the latest version of HAPI FHIR R4 [54].



9471307003
42 Unspecified
Lung carcinoma

Persist id: ABCD
Cancer details:
Lung carcinoma, stage II, pT1 pN0 M0 R0 G1
Diagnostic date: 21.10.2020

Treatment History
10.09.2020: Chemoterapy
01.09.2020: Chemoterapy
20.08.2020: Radiotherapy

Figure 3. Patient (**left**) and clinician (**right**) mHealth App first version.

4. Microservices to Support Fully Symmetric Interaction Model

4.1. End-to-End Multilingual Text-To-Speech Synthesis

The first microservice is Speech synthesis, the Text-to-speech (TTS) microservice. It mainly generates audio files with given transcriptions for the ECA that communicate with the patients. In short, the sequence-to-sequence model optimized for TTS is used to 'map' a sequence of letters to a sequence of phonemes. The TTS architecture used in [55] was developed for real-time or close to real-time systems by combination of two neural network models: a feature prediction NN model and a flow-based neural-network-vocoder WaveGlow. The model from [55] is outlined in Figure 4. It consists of an embedding layer that creates a 512-dimensional vector. The embedding vectors are directed into a series of three 1-D convolutional layers, each layer with 512 filters with length of 5. Each convolutional layer is followed with a mini-batch normalization and ReLU activation.

After the convolutional block, the tensors are fed to a bidirectional LSTMs and the feedback and backward results are concatenated. Since the decoder is implemented with a recurrent architecture, the outputs of the previous step ($i - 1$) are considered in each next step (i). The soft-attention mechanism represents a crucial element in this process. To create attention, the mechanism forms a context vector at each decoding step and updates the attention weight accordingly. The context vector (C_i), denoted by Equation (1), is computed as product of encoder output (h) and attention weight (α):

$$C_i = \sum_{j=1} \alpha_{ij} h_j \tag{1}$$

where the attention weight α_{ij} is calculated as:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1} \exp(e_{ik})} \tag{2}$$

where e_{ij} represents the energy and is calculated by a hybrid approach considering both location-based and content-based attention:

$$e_{i,j} = \omega^T \tan\mathrm{h}(W_{S_{i-1}} + Vh_j + Uf_{i,j} + b) \tag{3}$$

where S_{i-1} represents the previous state of the decoders LSTM, h_j represents the jth hidden encoder state, and $f_{i,j}$ location-signs calculated as a convolution operation f over the previous attention weight (α_{i-1}). W, V, U, and b are trained parameters.

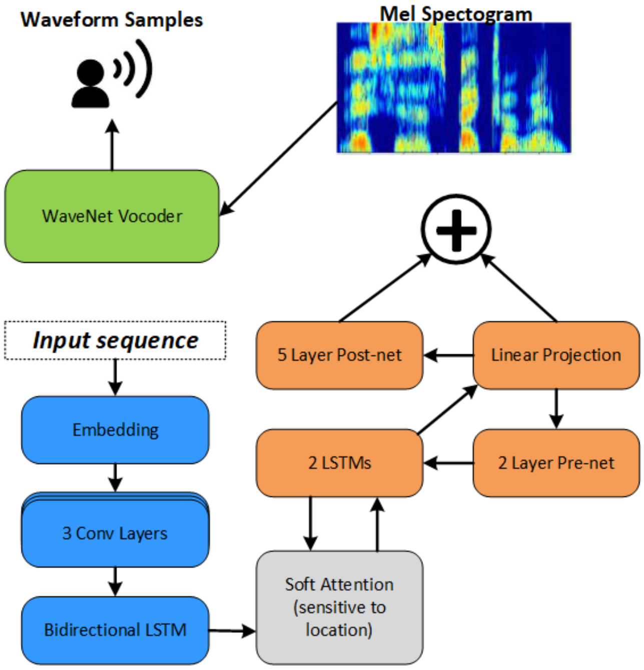


Figure 4. Tacatron 2 Model.

The output of the decoder is a predicted spectrogram. To improve the spectrogram quality, it is passed through the PostNet module; a stack of five one-dimensional convolutional layers with 512 filters in each one and a filter size of 5. Each layer (except the last one) is followed by batch-normalization and tangent activation. Finally, to transform the feature representation (i.e., spectrogram) into waveform (i.e., speech) a WaveNet architecture is used [56]. It consists of 30 dilated convolutional layers segmented into 3 cycles with the dilation rate of $2^{k \ (mod\ 10)}; where\ k\ \in [0, 30]$. To compute the logistic mixture distribution, the WaveNet output is finally passed through a ReLU activation, followed by linear projection. The used loss function is the negative log-likelihood.

4.2. End-to-End Multilingual Speech Recognition

Automatic speech recognition (ASR) represents the second microservice. It is implemented to support the spoken language interface in the Health App and feed the survivor's answers to the dialog management component (i.e., Rasa chatbot) where language is determined and processed. For the project PERSIST we deliver SPREAD, an E2E ASR system based on B × R Jasper model (Figure 5) [57], where B represents number of blocks, and Figure 4 represents the number of subblocks. Each subblock applies 1D convolutions, batch normalization, clipped ReLU activation, dropout, and residual connections. To improve training, we further introduce a new layer-wise optimizer called NovoGrad.

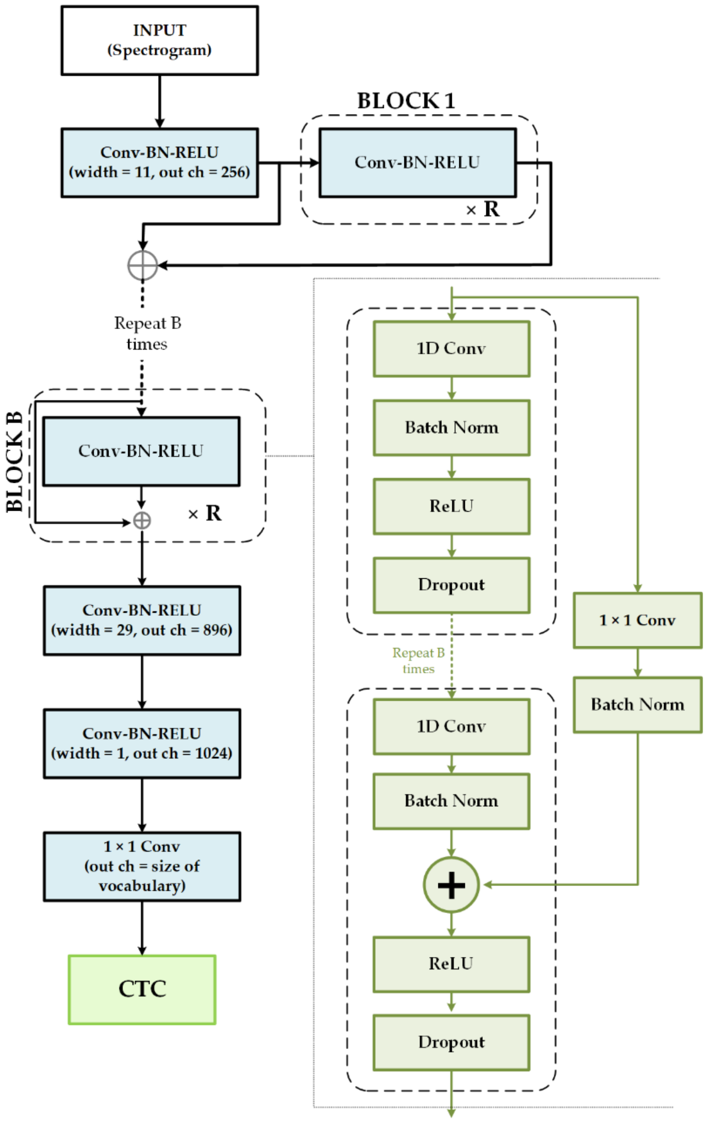


Figure 5. SPREAD ASR based on a B × R Jasper Model [57].

As highlighted in Figure 5, each residual connection is first projected through a 1 × 1 convolution. This enables the algorithm to account for different numbers of input and output channels. The residual connection is in this way added to the output of the last 1D-convolutional layer in the block before the clipped relu activation and dropout. Next, the output goes through a batch norm layer and the output of the batch norm layer is added to the output of the batch norm layer in the last sub-block. The overall sum is then passed through the activation function ReLU and dropout to produce the output of the current block B. All Jasper-based models have four generic convolutional blocks: one pre-processing, to reduce the time dimension of input speech signal, and three post-processing at the end. The first post-processing block performs a dilation of 2 to increase the model's receptive field while the last two post-processing blocks are fully connected. These are used to project the final output to a distribution over characters.

As outlined in Figure 5, a decoder based on Connectionist Temporal Classification (CTC) [58] is used to transform the output of the model in a sequence of letters corresponding to the speech input. In contrast to attention mechanism-based ASRs [59], the CTC decoder uses Markov assumptions to efficiently solve sequential problems by dynamic programming.

This allows the ASR to perform frame-by-frame label prediction with low computational cost by performing a greedy search and make it applicable even for long audio sequence. An E2E Inference is defined as classification of most problem grapheme sequence $\hat{W}$ in a given audio input X, i.e.,:

$$\hat{W} = \underset{W \in V}{argxmaxp}(W|X) \tag{4}$$

where $X = (x_1, \ldots, x_T)$ is a T-length speech feature sequence and $W = (w_1, \ldots, w_N)$ is an N-length grapheme sequence (i.e., a word sequence). Thus, at frame t, x_i is a D-dimensional speech feature vector and w_n is a word in vocabulary V on index n. The main problem of the ASR is therefore how to calculate the posterior distribution $p(W|X)$.

CTC formulation follows originates from Bayes decision theory and defines posterior distribution as:

$$p(C|X) = \sum_{Z} p(C|Z, X)p(Z|X) \tag{5}$$

$$\approx \sum_{Z} p(C|Z)p(Z|X) \tag{6}$$

CTC formulation uses L-length letter sequences, $C = (c_1, \ldots, c_L)$, with a set of distinct letters, U and a 'blank' symbol C' to denote the letter boundary and handle letter repetition:

$$C\prime = \{< s >, c_1, < s >, \ldots, < s >, c_L, < s >\} \tag{7}$$

where:

$$c_l = ' < \mathrm{s} >' \Big| \mathrm{U};\ l = 1,\ \ldots,\ 2L + 1 \tag{8}$$

which means that '<s>' is blank if *l* is odd and '<s>' is a letter from U if *l* is an even number. CTC also uses a conditional assumption $p(C|Z, X) \approx p(C|Z)$ to simplify the dependency between acoustic model and the letter models used in CTC.

4.3. *Embodied Conversational System and Embodied Conversational Agent*

A Rasa NLU [60] and ECA Framework [61] represent the final set of microservices and constitute an Embodied Conversational System which is capable to create responses in natural language as well as 'visualize' them. The Rasa Chatbot is used to manage the discourse between the survivor and the system. It is implemented as an API and the Rasa NLU represents the engine of the chatbot. The chatbot is running on a Linux server and is programmed in python and YAML language. The first version of our API implements 18 standardized patient reported outcomes (PROs) as storylines in six languages used in the PERSIST Clinical Study [62]. For storing the data, Rasa API uses the SQLite database which is possible with a function called SQLTrackerStore in the Rasa chatbot. POST and GET requests are used to store information, such as patients' answers, questionnaires, and all events that are triggered in a specific conversation.

The ECA Framework is designed to transform plain text sequences generated by the chatbot into multimodal responses incorporating gestures. The proprietary algorithm [61] is highlighted in Figure 6.

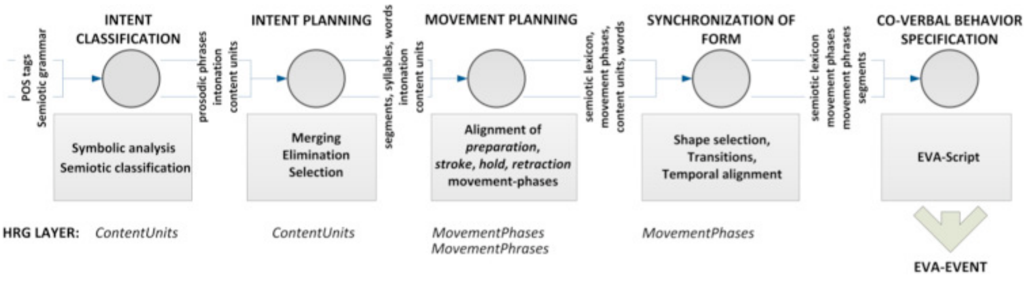


Figure 6. The algorithm for the generation of expressive co-verbal behavior.

It is based on the idea of segmenting a text sequence into non-verbal conversational intents and represent them as content units (CUs). As outlined in Figure 6 the algorithm performs the five phases in order to generate and synchronize no-verbal behavior with speech signal. In phase 1, the intent classification, it tries to recognize morphosyntactic patterns in the text and assign them and a communicative intent. The input of phase 1 is the POS-tagged text and the Semiotic Grammar; where the semiotic grammar represents a finite parametric space of the intent M_τ. The output of phase 1 is a set of multiple content units representing all possible communicative intents recognized in the text sequence. The content units identified may overlap and introduce inconsistencies. Thus, in phase 2, the intent planning, the algorithm needs to resolve these inconsistencies by elimination, margining and selection process. The algorithm considers prosodic alignment (i.e., prosodic phrases and intonation) to define final sequence of communicative intents to be 'visulalized'. In phase 3, the movement planning, for each planned intent an appropriate movement structure, i.e., the prosody of movement represented through movement phases (preparation, stroke, retraction, and hold), must be defined. Namely, a gesture phrase $\hat{G}$, visualizing the input text, is then defined as a sequence of gestures G_i visualizing each such content unit:

$$\hat{G} = (G_1, \ldots, G_N) \tag{9}$$

$$\hat{G} = H(CU_1, t_1) \times \ldots \times H(CU_N, t_N) \tag{10}$$

where each gesture G_1 represents a visualization of a specific CU with a movement model $\hat{H}$ and over time t. The operator × represents the successive execution of the movement models. Movement model $\hat{H}$ is 'sum' of animated sequence performed to visualized shapes/poses belonging to one of the movement phases executed over time t. Thus for each movement model $\hat{H}$ and 'end-pose' must be selected that can be viably animated given the duration of the movement phase. This is implemented in phase 4 of the algorithm, the synchronization of form. The selection of shape depends the type of movement-phase (inherently related to the supposed power of movement), the conversational intent and its possible 'semantic representation' and utterances (words or syllables) the non-verbal behavior is intended to visualize. Phase 4 adds shape to the structure the movement models.

In order to animate the gesture $\hat{G}$ movement models are finally transformed into a script, understandable to the ECA realization engine. This is carried out in final phase 5 of the algorithm entitled co-verbal behavior specification. We choose a proprietary EVA-Script notation s [61]. In the EVA-Script notation, each movement model $\hat{H}$ is formalized as simultaneous execution within the block <*bgesture*>. The Poses P within stroke phases and the preparation phases are represented as <*unit*> blocks within <*bgesture*> Each <*unit*> block contains the configuration of individual movement controllers involved in the representation of the pose. Since the hold phases and the retraction phases only represent the existing shape being withheld or retracted into the neutral state, they are added to the <*unit*> block in the form of attributes *DurationRetraction* and *DurationHold* in the block.

5. Results

For the preliminary evaluation, The PERSIST platform was deployed on two different physical servers at the University of Maribor, FERI. The detailed functionality of the system is highlighted by the Figure 7. As highlighted by Figure 7. The main actors of PERSIST are the clinician who defines and schedules an activity as part of patient's care workflow (i.e., phase 1 in Figure 7) and the patient, who executes the activity (i.e., phase 3 in Figure 7). The MSN and OHC represent the main services of the system. The OHC is used to store data and automate the execution of the clinical workflow. The MSN is used to implement activities and make their execution more natural by delivering the symmetric model of interaction.

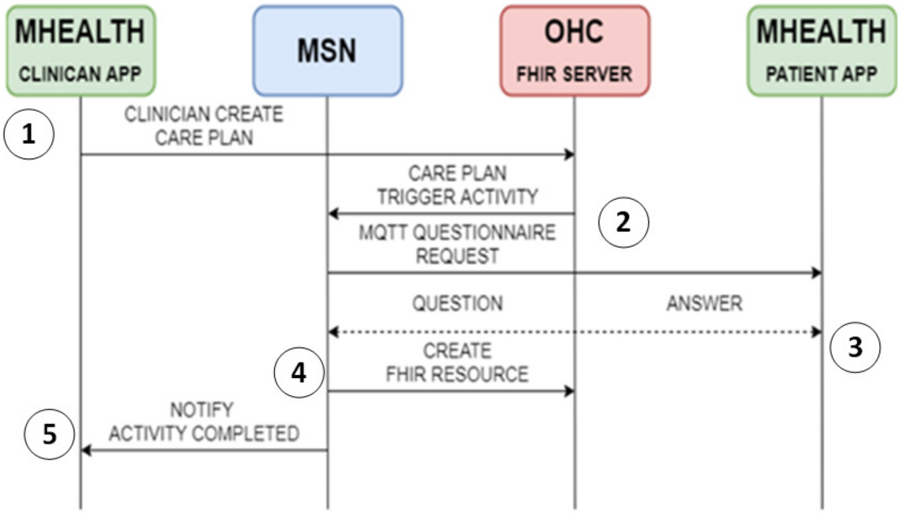

MHEALTH
CLINICAN APP
MSN
OHC
FHIR SERVER
MHEALTH
PATIENT APP
1
CLINICIAN CREATE CARE PLAN
CARE PLAN TRIGGER ACTIVITY
2
MQTT QUESTIONNAIRE REQUEST
QUESTION
ANSWER
3
4
CREATE FHIR RESOURCE
5
NOTIFY ACTIVITY COMPLETED

Figure 7. Integration of PROs into electronic health records as FHIR clinical resources: Functional Flow. The numbers 1–5 in the figure outline the phase of the integration process. 1—allocation of an activity, 2—request for execution of the activity, 3—implementation of the activity, 4—creation of resource and 5—completion of the activity.

As highlighted in Figure 7, the integration of a PRO starts in phase 1, when a clinician specifies a 'Care plan' FHIR resource (FHIR CarePLan: https://www.hl7.org/fhir/careplan.html, last visited 19 June 2021) for the patient. The care plans can be used for a general practitioner to schedule and keep track of when their patient is due to carry out a specific activity. In case of PERSIST, a self-report. Exploiting this resource, the OHC FHIR server is capable to automatically trigger a request for the 'todo'. This request is triggered by sending a 'notification' to the patient via the MQTT Broker hosted on the MSN (i.e., phase 2 in Figure 7). The notification about the triggered activity is sent through MSN's MQTT broker to the patient mHealth App in JSON data format. The patient can see five types of notifications: a request to fill the questionnaire, a request to provide information about the mood, a request to record a diary, a notification of a received message from the clinician, and other notifications. For all notification types, after the activity or 'todo' has been fulfilled the MSN automatically transforms the response into FHIR resource (phase 4 in Figure 7). Three types of resources are used, 'Observation' (FHIR Obsevation: https://www.hl7.org/fhir/observation.html, last visited 19 June 2021) to store reports regarding well-being (i.e., mood reports) and biometric data, 'DocumentReference' (FHIR DocumentReference: https://www.hl7.org/fhir/documentreference.html, last visited 19 June 2021) to store diary recordings and 'QuestionaryResponse' (FHIR QuestionarryResponse: https://www.hl7.org/fhir/questionnaireresponse.html, last visited 19 June 2021) The system also notifies the clinician that an activity was completed (phase 5 in Figure 7).

Figure 8 highlights the dataflow for the case of a request to answer pro (i.e., phase 3 in Figure 7). As highlighted in Figure 8, the main actor of the execution of the activity is the patient who uses the Health App to carry out the activity. The MSN and ECA (Chatbot) personalize interaction by delivering symmetric interaction on both input and output. The OHC implements JSON Web Tokens (RFC 7519) to ensure claims between two parties (i.e., mHealthApp and MSN) are secure. The process is initiated when a patient clicks on the notification request and thereby starts the question and answering (Q&A) dialog. In this exchange the ECA delivers questionnaires as multimodal responses and the user delivers the answer by clicking on the option or by answering with speech.

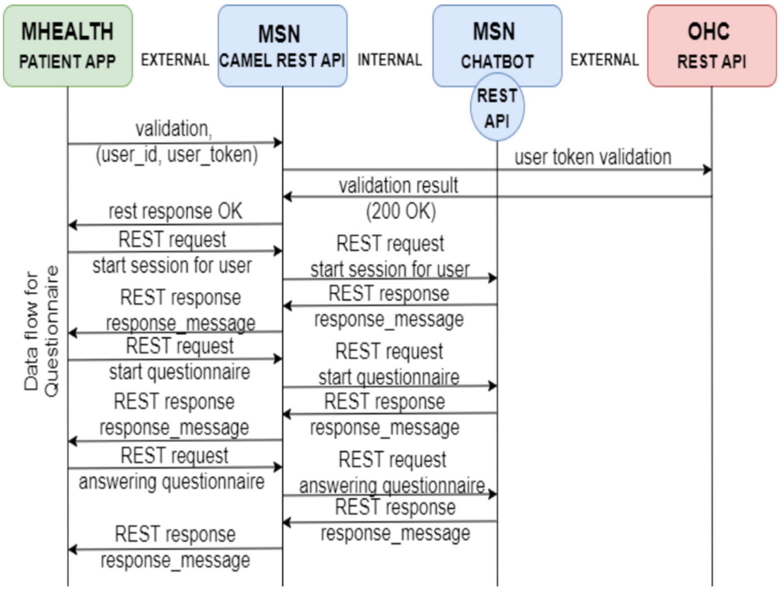


Figure 8. Functional flow of patient's executing an activity, i.e., answering a PRO.

As outlined in Figure 8, the system implements a symmetric interactive system in which users can answer questions from standardized questionnaires for the specific PRO. Questionnaires are available in the six different languages of the PERSIST project (Slovenian, English, Russian, Latvian, French, and Spanish). In order to, support multimodality on the output side, the system visualizes the chatbot generated information and represents them by the female agent 'Eva' and the male agent 'Adam' (Figure 9). Non-verbal elements are associated with speech. Unannotated texts are given as multimodal output which offers a spoken communication channel as well as a synchronized visual communication channel. BGM is synchronizing the verbal and non-verbal elements for our ECAs to act more naturally, more human-like. On the input side, the system accepts responses in text or speech format. To properly map the user response to the answers expected by PROs, a word-to-concept mapping is delivered as part of spoken language understanding.

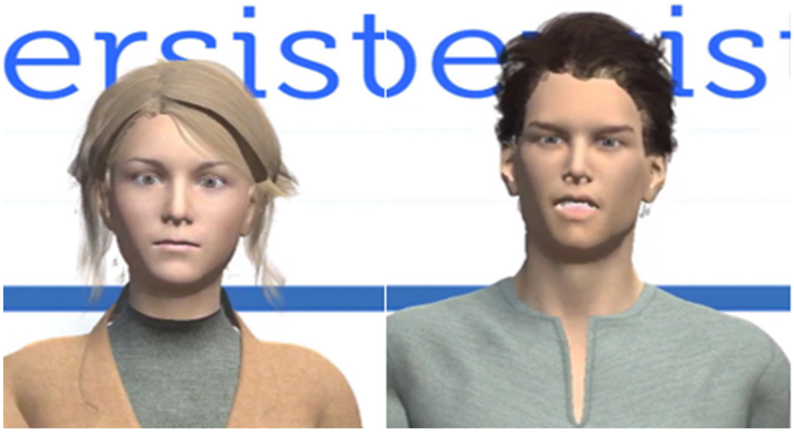


Figure 9. Visualizing conversational response with ECAs.

The proposed system was deployed on a server that hosts five virtual machines over the Proxmox VE 6.3-2 for the needs of the PERSIST project. That server is running the Xubuntu 20.04 LTS operating system. There are no virtual machines on the (other) server, named PERSIST_INFERENCE. On that server, running the Ubuntu Server 20.04 LTS OS, are microservices for ASR and TTS. TTS and ASR services are integrated using predefined topics, Kafka producers and consumers. Both microservices are being developed using

NVIDIA Triton Inference Server with TensorRT in Python programming language. The ECA microservice implementing the virtual agent is yet to be fully integrated and is operating as a standalone service. The specification for the infrastructure is lightweight with 8 GB of RAM for the Apache Kafka version 2.13-2.7.0 and the Apache Came version 3.4.0 with Apache ActiveMQ Artemis version 2.17.0. The 32 GB of RAM is provided to the Rasa chatbot version 2.1 which requires more memory capacity. Every building block has assigned 1 socket with 4 processor cores and 32 GB of SSD. To evaluate the hardware performance of the system, we simulated the load on the system by measuring CPU usage, memory usage, and average response time for both Camel and the Chatbot. The results are outlined in Figures 10–12.

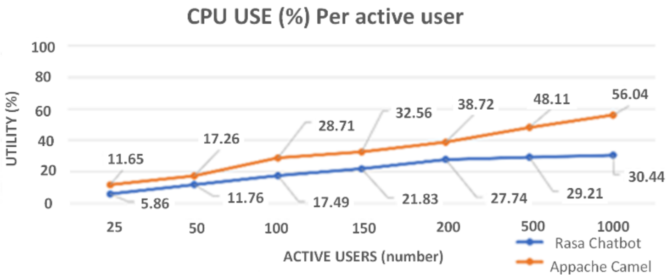


Figure 10. CPU use (%) per active users.

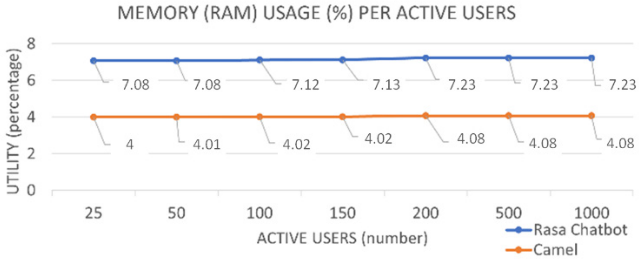


Figure 11. Memory consumption (GB) per active users.

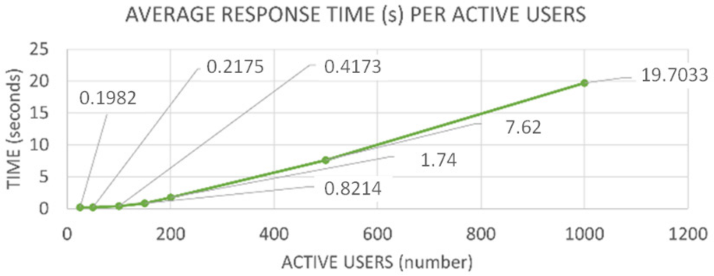


Figure 12. Graphical results of average response time per active user.

As can be seen from data in Figure 10, with the duplication of active users in tests, CPU usage is rising linearly from 11.65% on 25 active users, to 56.04% on 1000 active users for Camel, and also mostly linear from 5.86% on 25 active users, to 30.44% on 1000 active users for Rasa chatbot. In Figure 11, volatile memory was stagnating on both the Camel and the Rasa chatbot and proved independent of the increase of users. On Camel, memory usage was near 50%, and on the Rasa chatbot near 25%. Figure 12 shows MSN's internal average response time on requests between 25 and 1000 active users. At 25 active users, response time is 0.1982 s, and it is increasing linear as the number of users is increasing, up to the point of 200 active users where we see 1.74 s response time. From there it starts going up more exponentially to 197,033 s of delay when 1000 active users appear.

The preliminary evaluation of the end-to-end ASR system was delivered for the English language. We trained the models on DGX-1, 8 × V100, 8 × 32 g GPUMEM, and evaluated models on a workstation with 2 × RTX8000, 2 × 48 g GPUMEM. The audio dataset was roughly 1308.61 h. The best Jasper model reached 0.22 WER. The model was trained with *relu* encoder. To preliminarily evaluate the quality of end-to-end TTS, MUSHRA listening tests [63] were delivered for the English models among the PERSIST consortium partners. 21 consortium members participated. Overall, the best rated model was the model implemented as the Tacotron2 + Waveglow combination, represented in Section 4.2. It was evaluated with an average score of 75 on 100 level scale. The results show that speech generated by this model is intelligible and understandable, however, may sometimes sound machine-like. The evaluation of the multimodal conversational response was reported in [61]. 30 individuals assigned an average score of 3.45 on the 5-level Likert scale. This clearly implies that the system can produce a more viable and more believable user interface. To fully evaluate the ECA, responses in a targeted real-life environment with the same approach will be adapted in the project PERSIST. The patients will evaluate five dependent variables, describing the quality of the presentation on a 5-level Likert scale. In addition to gesture quality (e.g., form, dynamics, fluidity, synchronization), they will also report on how understandable the represented the content is (the sixth variable). After observing both instances (text + speech, and ECA with gestures), they will be asked to identify their general perception of the observed viability and human-likeness, expressed via the final, seventh dependent variable, which was also assessed on a 5-level Likert scale.

6. Discussion

However, the main challenges for wide adaptation of PGHD in clinical practice include usability (i.e., integration, interoperability with existing EHRs) and sustainable quality of results (i.e., patient motivation and adherence) [21,37] In order to address the interoperability and suitability of the collected resources we delivered a FHIR Methodology. The system presented includes patient/clinician mobile applications, an OHC FHIR server, and the MSN. The patient/clinician mobile applications are designed according to the security concerns and clinical trial requirements. Interoperability among the components is provided by the OHC FHIR server. OHC provides the framework and set of tools for the integration, ingestion, storage, indexing, and surfacing of patient information. It is an innovative, open digital integration hub with the proven ability to deliver the speed, scale, and flexibility needed to securely gain value through the integration of health systems. By consolidating data in a standardized format (FHIR) from across a broad range of systems of record, OHC enables innovation through near-real-time access to longitudinal patient records. The APIs provides opportunities to flexibly design services that can seamlessly ingest discrete data from the source (i.e., the EHR platform) into a third-party application (i.e., clinical or patient mHealthApp.) The proposed approach is further supported by recent trends in health care IT systems. Namely, FHIR has been recognized as an approach suitable for citizen developers since it also supports 'low-code/no-code' solutions [21]. This trend allows a user-centric design that creates intuitive data visualizations for transdisciplinary collaborations, including citizens, cognitive scientists, bioinformaticians, and clinicians [21]. For the data collected to be valuable, however, the whole IT system should be transformed to support the FHIR methodology. Several studies indeed report on the issue of actionability. If the PGHD is to drive the clinical decision workflow, the clinicians should be able to seamlessly exploit other data from the EHRs. Thus, our future efforts will be directed towards the transformation and ingestion of EHRs from existing IT platforms into FHIR ready server. Based on the preliminary studies, the main activities will involve the definition of an ontology (i.e., mapping model) that will correlate existing fields with specific FHIR resources. Moreover, since most of the information in existing EHRs is stored as partially structured or unstructured text, a specific focus will be directed towards extracting information using modern NLP techniques and data to concept mapping.

The other challenge related to the integration of PGHD relate to the patient's perspective; i.e., long-term sustainability and quality of collected information [36,37]. Familiarity, perceived complexity, and trustworthiness represent the main drivers of patient adherence [38]. To address this challenge, the MSN delivers a microservice infrastructure. In this infrastructure, the building blocks are each divided into their own service and are scaled so that the services are distributed among the servers and can be replicated if needed. A fully articulated ECA was deployed as the central technology to implement more natural human-machine interaction. The EVA framework is capable of capturing various contexts in the "data" and providing the basis to analytically investigate various multidimensional correlations among co-verbal behavior features. The role of the EVA realization framework is to transform the co-verbal descriptions contained in EVA events into articulated movement generated by the expressive virtual entity, e.g., to apply the EVA-Script language onto the articulated 3D model EVA in the form of animated movement [43]. Also, multilingual properties of our ECAs differentiates it from other similar ECAs. Trustworthiness is a clinical value which has a significant impact on adherence mitigating pervasive threats to health and wellbeing [64]. The applied model of symmetric multimodality for dialogue systems enables the ECAs to deliver and understand all-natural input/output modes, including speech, gestures, and facial expressions. This makes the synthetic interfaces significantly more familiar and trustworthy [38]. This is significant since trustworthiness is one of the building blocks of patient compliance and responsiveness [65]. The Chatbot API is using PREMs and PROMs to see the patients' health status and the patients' perceptions of their experience whilst receiving treatment. All sent and received data is in JSON format due to easy usage for transmitting data in web applications, and better representation and understanding. That is important because doctors should see the history of the report for specific patients in an understandable form in order to give more accurate decisions. For conversations between the Rasa Chatbot API and patients, it is important to configure story.yaml which contains the flow of the conversation or the order of the intents to be executed, which depends on the patients responses [66].

Overall, existing literature implies the consensus on the effects of an ECAs may have on patient adherence. Compared to chatbots and 2D agents, the fully articulated Embodied conversational agents have been proven to decrease the complexity of user interfaces and significantly contribute to familiarity and long-term sustainability [29]. Namely, fully articulated ECAs have a virtual body they can exploit to generate non-verbal cues with significant impact on understanding and cohesion of information exchange and the believability/trustworthiness of the digital entity. However, Ciechanowskiet et al. [67] note that the phenomenon off ''uncanny valley" may have significant negative impact on the overall user experience with articulated entities compared to ''disembodied" agents. Therefore, in our future efforts we will focus specifically on the synchronization of non-verbal behavior with speech. We plan to deliver a comparison study and Wizzard of Oz experiments to clearly define the design features of the symmetric model of interaction. Our preliminary experiments also showed that the 'quality' of synthesized behavior is closely related to hardware requirements. The model deployed for this study (Section 4.3) is not end-to-end and its computational requirements go well beyond any mobile or end-user device. As a result, our future activities will also investigate and end-to-end deployment of ECAs and non-verbal behavior generation models.

7. Conclusions

In this paper, we have represented a holistic approach towards sustainable collection of PGHD and PROs and their efficient integration into clinical workflow. PGHD may significantly contribute to personalized care and early identification related to psychological and physiological symptoms and negative health outcomes (e.g., cancer progression, toxicity, psychological distress). The system proposed in the paper represents an opportunity to integrate the possible benefits and deliver them to the patients. The system includes patient/clinician mobile applications, an OHC FHIR server, and a MSN. One of the major limitations stems from the inflexibility of existing healthcare platforms to adapt to FHIR or any other standardized. The lack of objective evaluation of the relevance and efficiency represents another limitation. Similarly, as related research, this study addresses the technologies from the prototype (proof-of-concept) perspective. The technology was evaluated on component basis, statistically, and on a short-term-use basis. However, within the project PERSIST, we plan to execute a 6-month final clinical evaluation with 160 cancer survivors and over 20 clinicians. The final limitation also arises from to the feasibility nature of the study and relates to usability. Although the co-creation activity implemented to design the system addressed the issues from the perspectives of multiple stakeholders, the clinical setting is limited to oncology and survivors of breast and colon cancer. Thus, results may not sufficiently reflect requirements of cancer patients (i.e., during treatment) or patients suffering from other (chronic) diseases.

Author Contributions: All authors equally contributed to the conceptualization, the layout of the research plan and the development and deployment of the system. All authors contributed to the writing and reviewing of the paper. Additionally, I.M. Lead the research and integration activities defined and reviewed the methodology, contributed the original draft, co-wrote, and revised the paper and the revision. V.Š., D.H., G.A., and R.P.L. contributed to the validation and co-wrote the original and revised versions of the paper. For the implementation of microservices, V.Š. oversaw the hardware and software infrastructure. U.A. oversaw the protocol. D.H. oversaw the delivery of chatbot. M.R. and I.M. oversaw delivering the symmetric model of interaction. M.R. oversaw ASR and I.M. oversaw delivering ECA service. G.A. oversaw the mHealthApp and R.P.L. oversaw the FHIR resources and OHC FHIR server. All authors have read and agreed to the published version of the manuscript.

Funding: This study is part of the project 'PERSIST: Patient-centered survivorship care plan after cancer treatment' that has received funding from the European Union's Horizon 2020 research and innovation program (GA No. 875406).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and as part of protocol registered at ISRCTN: https://doi.org/10.1186/ISRCTN97617326, accessed on 19 June 2021. The protocol was approved by the ethical committees in Belgium, Latvia, Slovenia, and Spain.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to restrictions apply to the availability of these data.

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. National Health Council. What are Clinician-Reported Outcomes (ClinROs)? National Health Council, 2019; Available online: https://nationalhealthcouncil.org/coa-series-what-are-clinician-reported-outcomes-clinros/ (accessed on 19 June 2021).
2. Health IT, Office of the National Coordinator for Health Information Technology (ONC), US Department of Health Human Services. What Are Patient-Generated Health Data? Available online: Healthit.gov/topic/otherhot-topics/what-are-patient-generated-health-data (accessed on 15 October 2019).
3. Fauzana, N.; Gulcharan, B.I.; Azhar, M.A.; Daud, H.; Mohd, N.N.; Taib, I. Integrating Emerging Network Technologies to Heart Rate Monitoring System to Investigate Transmission Stability and Accuracy: Preliminary Results. *Int. J. Electr. Eng. Comput. Sci. (EEACS)* **2021**, *3*, 21–22.
4. Sawssen, B.; Okba, T.; Noureeddine, L. A Mammographic Images Classification Technique via the Gaussian Radial Basis Kernel ELM and KPCA. *Int. J. Appl. Math. Comput. Sci. Syst. Eng.* **2020**, *2*, 92–98.
5. Zheng, Q.; Yang, L.; Zeng, B.; Li, J.; Guo, K.; Liang, Y.; Liao, G. Artificial intelligence performance in detecting tumor metastasis from medical radiology imaging: A systematic review and meta-analysis. *EClinicalMedicine* **2021**, *31*, 100669. [CrossRef]
6. Inès, A.; Zgaya, H.; Slim, H. Workflow tool to Model and simulate patients paths in Pediatric Emergency Department. *Int. J. Electr. Eng. Comput. Sci.* **2020**, *2*, 73–78.
7. Abdelnabi, M.L.R.; Jasim, M.W.; El-Bakry, H.M.; Taha, M.H.N.; Khalifa, N.E.M.; Loey, M. Breast and Colon Cancer Classification from Gene Expression Profiles Using Data Mining Techniques. *Symmetry* **2020**, *12*, 408. [CrossRef]
8. Austin, E.; LeRouge, C.; Hartzler, A.L.; Segal, C.; Lavallee, D.C. Capturing the patient voice: Implementing patient-reported outcomes across the health system. *Qual. Life Res.* **2020**, *29*, 347–355. [CrossRef]

9. Groccia, M.C.; Guido, R.; Conforti, D. Multi-Classifier Approaches for Supporting Clinical Decision Making. *Symmetry* **2020**, *12*, 699. [CrossRef]
10. Ellwood, P.M. Outcomes Management. *N. Engl. J. Med.* **1988**, *318*, 1549–1556. [CrossRef]
11. Tarlov, A.R.; Ware, J.E.; Greenfield, S.; Nelson, E.C.; Perrin, E.; Zubkoff, M. The Medical Outcomes Study: An application of methods for monitoring the results of medical care. *JAMA* **1989**, *262*, 925–930. [CrossRef]
12. Bielli, E.; Carminati, F.; La Capra, S.; Lina, M.; Brunelli, C.; Tamburini, M. A Wireless Health Outcomes Moni-toring System (WHOMS): Development and field testing with cancer patients using mobile phones. *BMC Med. Inform. Decis. Mak.* **2004**, *4*, 1–13. [CrossRef]
13. Tran, C.; Dicker, A.; Leiby, B.; Gressen, E.; Williams, N.; Jim, H. Utilizing digital health to collect electronic pa-tient-reported outcomes in prostate cancer: Single-arm pilot trial. *J. Med. Internet Res.* **2020**, *22*, e12689. [CrossRef]
14. Wright, A.A.; Raman, N.; Staples, P.; Schonholz, S.; Cronin, A.; Carlson, K.; Keating, N.L.; Onnela, J.-P. The HOPE Pilot Study: Harnessing Patient-Reported Outcomes and Biometric Data to Enhance Cancer Care. *JCO Clin. Cancer Inform.* **2018**, *2*, 1–12. [CrossRef]
15. Rajguru, P.; Ryan, S.; McLaurin, E.; Wirta, D.; Grieco, J. A novel method for collecting patient reported outcomes (PROs): Developing and validating electronic PROs on a mobile smartphone platform. *Invest. Ophthalmol. Vis. Sci.* **2020**, *7*, 110.
16. Van Egdom, L.S.E.; Pusic, A.; Verhoef, C.; Hazelzet, J.A.; Koppert, L.B. Machine learning with PROs in breast cancer surgery; caution: Collecting PROs at baseline is crucial. *Breast J.* **2020**, *26*, 1213–1215. [CrossRef]
17. Kramer, L.L.; Ter Stal, S.; Mulder, B.; De Vet, E.; Van Velsen, L. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, e14058. [CrossRef]
18. Queirós, A.; Dias, A.; Silva, A.G.; Rocha, N.P. Ambient assisted living and health-related out-comes-A systematic literature review. *Informatics* **2014**, *4*, 19. [CrossRef]
19. Alosaimi, W.; Ansari, T.J.; Alharbi, A.; Alyami, H.; Seh, A.; Pandey, A.; Agrawal, A.; Khan, R. Evaluating the Impact of Different Symmetrical Models of Ambient Assisted Living Systems. *Symmetry* **2021**, *13*, 450. [CrossRef]
20. Laranjo, L.; Dunn, A.; 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. Inform. Assoc.* **2018**, *25*, 1248–1258. [CrossRef] [PubMed]
21. Jim, H.S.L.; Hoogland, A.; Brownstein, N.C.; Barata, A.; Dicker, A.P.; Knoop, H.; Gonzalez, B.D.; Perkins, R.; Rollison, D.; Gilbert, S.M.; et al. Innovations in research and clinical care using patient-generated health data. *CA Cancer J. Clin.* **2020**, *70*, 182–199. [CrossRef]
22. Rehman, A.; Naz, S.; Razzak, I. Leveraging big data analytics in healthcare enhancement: Trends, challenges and opportunities. *Multimedia Syst.* **2021**, 1–33. [CrossRef]
23. Resourcelist—FHIR v4.0.1. Available online: http://hl7.org/fhir/resourcelist.html (accessed on 1 April 2021).
24. Wald, J.S.; Sands, D.Z. Transforming Health Care Delivery Through Consumer Engagement, Health Data Transparency, and Patient-Generated Health Information. *Yearb. Med. Inform.* **2014**, *23*, 170–176. [CrossRef]
25. Shawar, B.; Atwell, E. Chatbots: Are they Really Useful? *LDV Forum* **2007**, *22*, 29–49.
26. Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM* **1966**, *9*, 36–45. [CrossRef]
27. Sharma, R.K.; Center, N.I. An Analytical Study and Review of open source Chatbot framework, Rasa. *Int. J. Eng. Res.* **2020**, *9*, 060723. [CrossRef]
28. Palanica, A.; Flaschner, P.; Thommandram, A.; Li, M.; Fossat, Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J. Med. Internet Res.* **2019**, *21*, e12887. [CrossRef] [PubMed]
29. Bibault, J.-E.; Chaix, B.; Nectoux, P.; Pienkowski, A.; Guillemasé, A.; Brouard, B. Healthcare ex Machina: Are conversational agents ready for prime time in oncology? *Clin. Transl. Radiat. Oncol.* **2019**, *16*, 55–59. [CrossRef]
30. Owens, O.L.; Felder, T.; Tavakoli, A.S.; Revels, A.A.; Friedman, D.B.; Hughes-Halbert, C.; Hébert, J.R. Evaluation of a Computer-Based Decision Aid for Promoting Informed Prostate Cancer Screening Decisions Among African American Men: iDecide. *Am. J. Health Promot.* **2019**, *33*, 267–278. [CrossRef] [PubMed]
31. Fitzpatrick, K.K.; Darcy, A.; Vierhile, M. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment. Health* **2017**, *4*, e19. [CrossRef]
32. Inkster, B.; Sarda, S.; Subramanian, V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR mHealth uHealth* **2018**, *6*, e12106. [CrossRef]
33. 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]
34. Gardiner, P.M.; McCue, K.D.; Negash, L.M.; Cheng, T.; White, L.F.; Yinusa-Nyahkoon, L.; Jack, B.W.; Bickmore, T.W. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Educ. Couns.* **2017**, *100*, 1720–1729. [CrossRef] [PubMed]
35. Girgi, A.; Durcinoska, I.; Levesque, J.V.; Gerges, M.; Sandell, T.; Arnold, A.; Delaney, G.P. The PROMPT-Care Program Group eHealth System for Collecting and Utilizing Patient Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care) Among Cancer Patients: Mixed Methods Approach to Evaluate Feasibility and Acceptability. *J. Med. Internet Res.* **2017**, *19*, e330. [CrossRef] [PubMed]

36. Kneuertz, P.J.; Jagadesh, N.; Perkins, A.; Fitzgerald, M.; Moffatt-Bruce, S.D.; Merritt, R.E.; D'Souza, D.M. Improving patient engagement, adherence, and satisfaction in lung cancer surgery with implementation of a mobile device platform for patient reported outcomes. *J. Thorac. Dis.* **2020**, *12*, 6883–6891. [CrossRef] [PubMed]
37. Tellols, D.; Lopez-Sanchez, M.; Rodríguez, I.; Almajano, P.; Puig, A. Enhancing sentient embodied conversational agents with machine learning. *Pattern Recognit. Lett.* **2020**, *129*, 317–323. [CrossRef]
38. Martin, L.R.; Williams, S.L.; Haskard, K.B.; DiMatteo, M.R. The challenge of patient adherence. *Ther. Clin. Risk Manag.* **2005**, *1*, 189–199. [PubMed]
39. Isbister, K.; Doyle, P. The blind men and the elephant revisited evaluating interdisciplinary ECA research. In *From Brows to Trust Evaluating Embodied Conversational Agents*; Ruttkay, Z., Pelachaud, C., Eds.; Springer: Dordrecht, The Netherlands, 2004; pp. 3–26.
40. Bickmore, T.; Gruber, A.; Picard, R. Establishing the computer–patient working alliance in automated health behavior change interventions. *Patient Educ. Couns.* **2005**, *59*, 21–30. [CrossRef]
41. Klaassen, R.; Bul, K.C.M.; Akker, R.O.D.; Van Der Burg, G.J.; Kato, P.M.; Di Bitonto, P. Design and Evaluation of a Pervasive Coaching and Gamification Platform for Young Diabetes Patients. *Sensors* **2018**, *18*, 402. [CrossRef] [PubMed]
42. Provoost, S.; Lau, H.M.; Ruwaard, J.; Riper, H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *J. Med. Internet Res.* **2017**, *19*, e151. [CrossRef]
43. Rojc, M.; Kačič, Z.; Mlakar, I. Advanced Content and Interface Personalization through Conversational Behavior and Affective Embodied Conversational Agents. In *Artificial Intelligence Emerging Trends and Applications*; Fernandez, M.A.A., Ed.; IntechOpen: London, UK, 2018. [CrossRef]
44. Brinkman, W.P. Virtual health agents for behavior change: Research perspectives and directions. In Proceedings of the Workshop on Graphical and Robotic Embodied Agents for Therapeutic Systems, Institute for Creative Technologies, USC, Los Angeles, CA, USA, 20 September 2016.
45. Stal, S.; Kramer, L.L.; Tabak, M.; Akker, H.O.D.; Hermens, H. Design Features of Embodied Conversational Agents in eHealth: A Literature Review. *Int. J. Hum. Comput. Stud.* **2020**, *138*, 102409. [CrossRef]
46. Friederichs, S.; Bolman, C.; Oenema, A.; Guyaux, J.; Lechner, L. Motivational Interviewing in a Web-Based Physical Activity Intervention with an Avatar: Randomized Controlled Trial. *J. Med. Internet Res.* **2014**, *16*, e48. [CrossRef]
47. Bickmore, T.W.; Caruso, L.; Clough-Gorr, K.; Heeren, T. 'It's just like you talk to a friend' relational agents for older adults. *Interact. Comput.* **2005**, *17*, 711–735. [CrossRef]
48. Ellis, T.; Latham, N.K.; DeAngelis, T.R.; Thomas, C.A.; Saint-Hilaire, M.; Bickmore, T.W. Feasibility of a Virtual Exercise Coach to Promote Walking in Community-Dwelling Persons with Parkinson Disease. *Am. J. Phys. Med. Rehabil.* **2013**, *92*, 472–485. [CrossRef]
49. Henkemans, B.O.A.; van der Boog, P.J.; Lindenberg, J.; van der Mast, C.A.; Neerincx, M.A.; Zwetsloot-Schonk, B.J. An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. *Technol. Health Care* **2009**, *17*, 253–267. [CrossRef] [PubMed]
50. Bickmore, T.W.; Schulman, D.; Sidner, C. Automated interventions for multiple health behaviors using conversational agents. *Patient Educ. Couns.* **2013**, *92*, 142–148. [CrossRef]
51. Sillice, A.M.; Morokoff, P.J.; Ferszt, G.; Bickmore, T.; Bock, B.C.; Lantini, R.; Velicer, W.F. Using Relational Agents to Promote Exercise and Sun Protection: Assessment of Participants' Experiences with Two Interventions. *J. Med. Internet Res.* **2018**, *20*, e48. [CrossRef]
52. Benze, G.; Nauck, F.; Alt-Epping, B.; Gianni, G.; Bauknecht, T.; Ettl, J.; Munte, A.; Kretzschmar, L.; Gaertner, J. PROutine: A feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer. *Ann. Palliat. Med.* **2019**, *8*, 104–111. [CrossRef] [PubMed]
53. Sayeed, R.; Gottlieb, D.; Mandl, K.D. SMART Markers: Collecting patient-generated health data as a standardized property of health information technology. *NPJ Digit. Med.* **2020**, *3*, 1–8. [CrossRef]
54. Versions—FHIR v4.0.1. Available online: https://www.hl7.org/fhir/versions.html (accessed on 1 April 2021).
55. Shen, J.; Pang, R.; Weiss, R.J.; Schuster, M.; Jaitly, N.; Yang, Z.; Wu, Y. Natural tts synthesis by condi-tioning wavenet on mel spectrogram predictions. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4779–4783.
56. Oord, A.V.D.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kavukcuoglu, K. Wavenet: A generative model for raw audio. *arXiv* **2016**, arXiv:1609.03499.
57. Li, J.; Lavrukhin, V.; Ginsburg, B.; Leary, R.; Kuchaiev, O.; Cohen, J.M.; Nguyen, H.; Gadde, R.T. Jasper: An End-to-End Convolutional Neural Acoustic Model. *arXiv* **2019**, arXiv:1904.03288.
58. Graves, A.; Jaitly, N. Towards end-to-end speech recognition with recurrent neural networks. In Proceedings of the International Conference on Machine Learning (ICML 2014), Beijing, China, 21–26 June 2014; pp. 1764–1772.
59. Chorowski, J.; Bahdanau, D.; Cho, K.; Bengio, Y. End-to-end continuous speech recognition using attention-based recurrent nn: First results. *arXiv* **2014**, arXiv:1412.1602.
60. Bocklisch, T.; Faulkner, J.; Pawlowski, N.; Nichol, A. Rasa: Open source language understanding and dialogue management. *arXiv* **2017**, arXiv:1712.05181.
61. Rojc, M.; Mlakar, I.; Kačič, Z. The TTS-driven affective embodied conversational agent EVA, based on a novel conversational-behavior generation algorithm. *Eng. Appl. Artif. Intell.* **2017**, *57*, 80–104. [CrossRef]

62. Mlakar, I.; Smrke, U. Clinical Study to Assess the Outcomes of a Patient-Centred Survivorship Care Plan Enhanced with Big Data and Artificial Intelligence Technologies 2021. Available online: https://www.isrctn.com/ISRCTN97617326 (accessed on 19 June 2021).
63. Schoeffler, M.; Bartoschek, S.; Stöter, F.R.; Roess, M.; Westphal, S.; Edler, B.; Herre, J. Web MUSHRA—A comprehensive framework for web-based listening tests. *J. Open Res. Softw.* **2016**, *6*. [CrossRef]
64. H2020 Project PERSIST. Available online: https://projectpersist.com/ (accessed on 31 May 2020).
65. Sofer, C.; Dotsch, R.; Wigboldus, D.H.; Todorov, A. What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychol. Sci.* **2015**, *26*, 39–47. [CrossRef] [PubMed]
66. Singh, A.; Ramasubramanian, K.; Shivam, S. Introduction to Microsoft Bot, RASA, and Google Dialogflow. In *Building an Enterprise Chatbot: Work with Protected Enterprise Data Using Open Source Frameworks*; Apress: Berkeley, CA, USA, 2019; pp. 281–302. [CrossRef]
67. Ciechanowski, L.; Przegalinska, A.; Magnuski, M.; Gloor, P. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Gener. Comput. Syst.* **2018**, *92*, 539–548. [CrossRef]