ICIS 2019 Proceedings | IS in Healthcare

FeelFit – Design and Evaluation of a Conversational Agent to Enhance Health Awareness

Pascal Meier
University of Osnabrück, pascal.meier@uni-osnabrueck.de

Jan Heinrich Beinke
University of Osnabrück, jan.beinke@uni-osnabrueck.de

Christian Fitte
University of Osnabrück, christian.fitte@uni-osnabrueck.de

Alina Behne
University of Osnabrück, alina.behne@uni-osnabrueck.de

Frank Teuteberg
University of Osnabrück, frank.teuteberg@uni-osnabrueck.de

FeelFit – Design and Evaluation of a Conversational Agent to Enhance Health Awareness

Completed Research Paper

Pascal Meier[1]
pascal.meier@uni-osnabrueck.de

Jan Heinrich Beinke[1]
jan.beinke@uni-osnabrueck.de

Christian Fitte[1]
christian.fitte@uni-osnabrueck.de

Alina Behne[1]
alina.behne@uni-osnabrueck.de

Frank Teuteberg[1]
frank.teuteberg@uni-osnabrueck.de

[1]Osnabrück University
Katharinenstr. 1, 49074 Osnabrück,
Germany

Abstract

In the course of digitalisation, healthcare systems are undergoing a major transformation. The generation and processing of health-related data are intended to improve health concerns. However, individual health awareness remains inadequate. To counteract this problem, issues in the fields of health awareness, wearable health monitoring systems, conversational agents, and user interface design were identified. Meta-requirements were derived from these issues and then converted into design principles. We developed the FeelFit conversational agent under consideration of those design principles. FeelFit measures vital parameters with various wearable sensors and presents them, enriched with personalised health information, to the user in the form of a conversation via individually configurable input and output devices. The conversational agent was evaluated by two experiments with 90 participants and a workshop. The results confirm a positive usability and task fulfilment of our conversational agent. Compared to known applications, the participants highlighted the more natural interaction and seamless integration of various sensors as strengths of FeelFit.

Keywords: Conversational Agent, Health Awareness, Wearable Health Monitoring System

Introduction

Conversational agents (CAs), including text-based chatbots and virtual private assistants (VPAs) such as Amazon Alexa and Google Home, have become increasingly widespread in recent years because of advances in spoken language processing, artificial intelligence, and conversation interface design (Luger and Sellen 2016). Especially in the smart home sector, CAs are used to respond to queries (such as weather reports) or commands (e.g., lighting control) using natural language. These spoken or written inputs make operation easier for users, since they can express concerns directly and do not have to adapt the intention to a conventional graphical user interface (Seeger et al. 2018).

At the same time, the spread of digital health services as well as wearable devices for self-monitoring that support a user's health awareness is observed (Watson 2016). These technologies motivate users to increase their activity levels and enable the self-management of chronic diseases. An insufficient understanding of one's own health poses major challenges (Prugger et al. 2011). A lack of awareness of one's blood pressure, blood glucose, weight, pulse, and movement, for example, can lead to the development of serious chronic illnesses (e.g., diabetes, chronic heart failure, hypertension) and hence higher costs for the healthcare system (Palumbo 2017). By using vital parameter monitoring systems in everyday life, costs savings up to 40% can be achieved (Paré et al. 2013). This is primarily because of reduced hospitalisations, which are also shortened and less severe (Paré et al. 2013). Furthermore, a regular overview of the above-mentioned vital parameters could underpin the benefits of drug therapies and thus motivate patients to adhere to their therapy plans. A substantial proportion of therapy costs result from insufficient patient adherence to therapy (Blaschke et al. 2012). One reason for this is a lack of immediate visibility of therapy effects, as reflected in improvements in the vital parameters. To support these improvements, wearable health monitoring systems (WHMSs) use sensors to send measured data to processing units, which process the data and present the findings to the user (Pantelopoulos and Bourbakis 2010). To date, WHMSs are primarily employed for fitness and lifestyle purposes because of the insufficient accuracy of sensors for medical application. However, the rising dissemination of fitness trackers and the quantified self community increase users' health awareness and support them in changing their personal lifestyle (Swan 2009). Improvements in the accuracy of fitness sensors may enable their use for medical purposes. For example, a recent study with over 400,000 users of the Apple watch has demonstrated that the device is able to detect atrial fibrillation by irregular heart rhythms (Stanford University School of Medicine 2019).

Despite the high potential of WHMSs, most manufacturers only support their own vital parameter sensors and do not offer the option of integrating data from other manufacturers' sensors. In addition, almost no WHMS includes a CA that enables a more natural user interaction. To address this, we developed a system named 'FeelFit' in a one-year research project. FeelFit utilises sensors from various manufacturers to monitor users' vital parameters, consolidates and integrates the sensor data, and outputs them via a conversational interface (CI). It is built as a modular system, which contains different components and thus enables a user-centred configuration. The components available in the current system include natural language input and output of data, display options on a smart mirror, and interaction with other mobile devices.

Related work has revealed that the scientific research on WHMSs and CAs in the context of health awareness still requires further investigation. In the field of information systems, Diederich et al. (2019) analysed the current research on CAs and identified, among other things, the need to formulate design principles (DPs) as well as to design and evaluate their instantiations. In addition, Gnewuch et al. (2017) designed a CA that increased the quality of customer service. The basis for this consisted of the collected meta-requirements (MRs) and the DPs for social CAs. Moreover, Bickmore et al. (2016) designed a conversational search engine interface that allowed people with low health and computer literacy to access information about clinical trials. The study demonstrated that all participants were more satisfied with the CI than with a conventional web form-based interface. However, it was not possible to formulate the search query directly and verbally. In another study, Gnewuch et al. (2018) developed literature-based DPs for CAs with a focus on providing energy feedback to consumers to increase energy use awareness. Despite these examples of designing CAs, no previous work could be identified that (a) integrates different sensors to monitor vital parameters, (b) offers the possibility of querying these, (c) outputs the data by a CA, (d) displays them on various devices to keep the user's inhibition threshold as low as possible, and (e) evaluates the DPs used for the artefact development. In view of this research gap, we derive the following research question (RQ):

RQ: *How can conversational agents for wearable health monitoring systems be designed and implemented to enhance users' health awareness?*

To answer this research question, we analysed pre-existing conversational applications in the field of health awareness. The findings are presented in section 2. Thereafter, in section 3, we describe the multi-methodological approach embedded in the design science paradigm. In section 4, the concept of using a CA in a WHMS is designed based on the developed DPs. This is followed by a presentation of the FeelFit application in section 5 and an evaluation of the FeelFit artefact in section 6. We then discuss our findings and implications in section 7 based on the evaluation. Finally, we name the limitations of our work and suggest future research opportunities.

Conversational Agents for Health Awareness

With the emergence of mobile devices, especially smartphones, new possibilities have developed in the healthcare sector for monitoring patients' health. Users are consequently able to document symptoms as they occur. For example, patients can write a diary about the course of their illness or note measured values (Mattila et al. 2008). Using wearable devices to measure vital parameters, the documentation of these values can be automated, with continuous monitoring made possible for patients (Pantelopoulos and Bourbakis 2010). Thresholds set by doctors or nurses (e.g. blood sugar levels between 72 mg/dL and 140 mg/dL) can also initiate notifications to the user if the vital parameters fall outside of the recommended range, thus creating a greater awareness of health.

Moreover, the movement of the quantified self has become popular in the healthcare sector in recent years. The meaning of the term 'quantified self' is to gain new knowledge of oneself by tracking one's own vital parameters to monitor, evaluate, and optimise one's own behaviour and health condition also through the use of social media (Baumgart and Wiewiorra 2016). Wearables that support a healthy lifestyle, such as smart watches, fitness or food tracker applications, and activity trackers with various sensors, are becoming increasingly popular, favoured because of the decreasing sensor size and the increasing use and user-friendliness of information technology. In this way, data can be collected, and findings about the user's fitness and health status can be summarised and visualised to provide him or her with relevant information (Lupton 2016).

Given developments in recent years in the fields of artificial intelligence, natural language processing, and CI design, CAs (e.g., chatbots and VPAs) have appeared as a stable platform for language-based interactions (McTear et al. 2016). To make the input and output of these devices intuitive for the user, CIs are used in embodied CAs and chatbots (Klopfenstein et al. 2017). Accompanied by the market maturity of chatbots and VPAs, the published research in the field of CAs has increased in the past three years (Diederich et al. 2019). In addition to the general research on user experience and the effects of CAs, domain-specific research has been conducted to demonstrate how CAs can be utilised in specific domains such as marketing and sales (Maass et al. 2011), customer service (Gnewuch et al. 2017), and interviewing (Nunamaker et al. 2011). Further studies have monitored the use of CAs in 'smart homes' and for bookings and orders, while others have dealt with their use in the health sector. Conversational interfaces are particularly common in the field of mental health, wherein they support psychotherapy in the treatment of depression and anxiety (Fitzpatrick et al. 2017), improve social skills for people with autism (Tanaka et al. 2017), and inform patients about their disease in order to train behaviours (Hudlicka 2013). Furthermore, there are CAs for the self-monitoring of asthma (Rhee et al. 2014) and for the telemonitoring of diabetes (Griol et al. 2013) and hypertension (Giorgino et al. 2005). In addition to scientific studies, companies such as ADA (ADA 2019), which provides conversational diagnoses, and DOCYET (DOCYET 2019), which refers patients to a nearby physician, have arisen and achieved success in the form of funding. In April 2019, Amazon announced the first HIPAA-compliant applications for Alexa. Apart from their application for delivery status checks for prescription orders, tracking of health incentive plans as well as communication and appointment agreements with healthcare providers, an application by Livongo allows users to manage their blood sugar levels by providing information about the readings, measurement trends, and health nudges (Jiang 2019).

However, to the best of our knowledge, no application was identified that provides continuously measured data from various sensors via a CA. Rather, the information with which CAs work based only on general information or perceived symptoms; the continuously measured data provided by WHMSs are not yet included. A WHMS combined with a CA allows users to check their health status at any time. Furthermore, enriched with educational information, the system simplifies users' access to their health data. The CA thus offers a supplement to professional treatment and examination by physicians, pharmacists, and nurses. For the system to provide users with consistent information on their health at all times, the individual subsystems must be co-ordinated to achieve a consistent user experience.

Multi-methodological Approach

As mentioned in section 2, several health applications already exist on the market. However, to date, no CA currently exists that includes vital parameters from different sources and service providers independent of the interaction device. This research gap has motivated us to develop an integrated conversational health agent. To achieve this, we applied the design science research paradigm (Hevner et al. 2004). Continuous design and evaluation cycles, in constant exchange with the knowledge base and the environment, enabled us to ensure a relevant and rigorous artefact development. Since designing a CA to provide vital parameters to the user is not sufficient to improve his or her health awareness, the developed CA was designed to be superior to existing fitness and health applications with regard to being informed about one's own health in everyday life. The artefact development was based on the design science research methodology by Peffers et al. (2007) (cf. Figure 1). In total, three iterations were performed until the final artefact was presented.

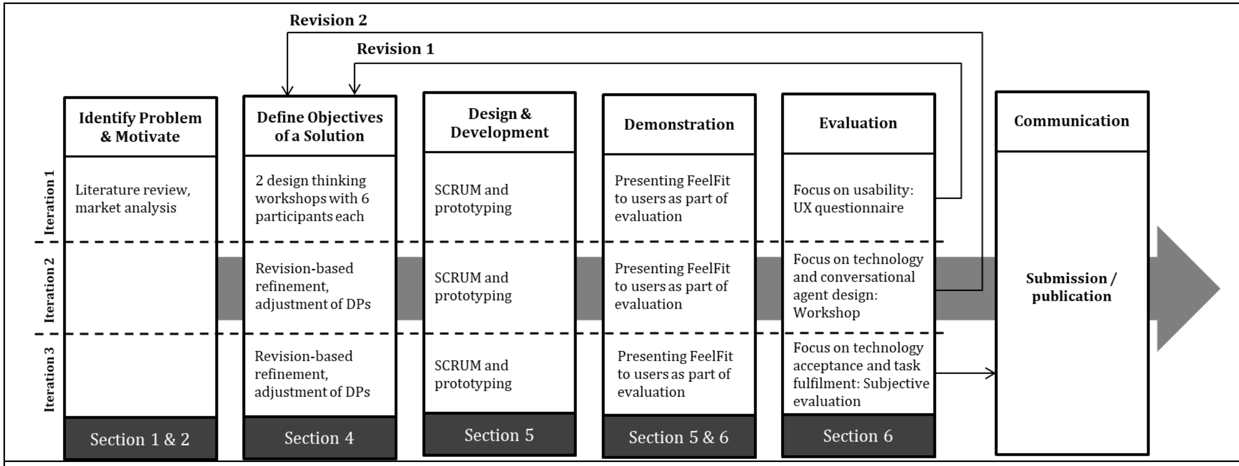


Figure 1. Design Science-based Research Approach

We examined the status quo by means of a structured literature analysis (vom Brocke et al. 2009) and a market analysis. The latter focused on the Apple App Store and the Google Play Store for mobile applications and on the Amazon Alexa and Google Assistant marketplaces for CAs. We combined the term *'conversational agent'* with the terms *'smart health', 'eHealth', 'mHealth', 'wearable health monitoring system', 'chronic disease', 'rehabilitation' 'diabetes', 'congestive heart failure', 'hypertension', 'fitness',* and *'nutrition'*. This led to 14 research papers and 11 applications relevant for our project. Furthermore, a concept matrix, as described by Webster and Watson (2002), helped to identify nine major issues and barriers for mobile health applications and CAs in the literature along with best practices. In the second step, we collected structured MRs for our CA. The identified issues were used to derive MRs in two design thinking workshops, each of which contained six participants and was conducted with researchers and practitioners from the fields of computer sciences and information systems, healthcare, and smart homes in October 2018. The workshops were recorded on tape, transcribed, and qualitatively evaluated (Mayring 2010). For the analysis, the transcribed text passages were coded with the software MaxQDA. To reduce potential subjective distortions, the coding was performed by two people independently of each other, and in the case of a different coding, the respective passages were discussed in the research team, and a final decision was made. We used Krippendorff's alpha (workshop 1: 0.82; workshop 2: 0.85) as the quality measure for the inter-coder agreement, which is above the 'customarily' required value of 0.8 (Krippendorff 2004).

We applied the design thinking method and consulted a total of 12 experts (Plattner et al. 2009). The MRs were consolidated in four DPs, which served as guidelines for the development process. Both MRs and DPs were further developed and refined in the following two iterations. The development process began with a definition of the concept of the FeelFit application. Based on this concept, we then set up the architecture for our system. In addition, we created prototypes of the different application interfaces (Dey et al. 2001). In this stage, we designed the mobile application with different tiles for each vital parameter, created exemplary conversation flows for the CA, and designed the visualisation on a smart mirror. The scope of included parameters and sensors can be configured individually in the mobile application. During the development phase, we used the SCRUM method to organise the project and realised the prototypes for the CA, mobile application, and smart mirror. A total of three development phases were carried out, each of which concluded with an evaluation to incorporate the findings into the improvement of the CA, the application, and the smart mirror. After the first iteration, we conducted a scenario-based evaluation with 54 participants, focusing on the usability of our artefacts. The second evaluation method was a workshop with six experts from the field of information technology and CAs. In addition to the improvement of the artefact, the developed MRs and DPs were also discussed and refined. As the final evaluation method after the third development phase, a survey with 36 participants was conducted to examine the task fulfilment and technology acceptance. The evaluation design and the collected results are presented in section 6.

Issues, Meta-Requirements, and Design Principles for a Conversational Agent Supporting Health Awareness

Derived from the issues (Is) identified in the literature and the market review, the MRs were elaborated in several workshops and refined by the last two iterations. The DPs were developed based on the MRs, and they were taken into account during the development of the prototype. The results are categorised into four areas: health awareness, CA, user interface design, and information architecture.

As discussed in section 2, the use of mobile technology in health is increasing, alongside progress in digitisation. However, users' health awareness remains inadequate (**I1**) (Kim and Xie 2017). In addition to the barrier of computer literacy, many users have limited access to health information and difficulties with interpreting this information (Mather and Cummings 2015). Since the CA aims to provide users with health-related data, it is important that the agent also provides educational health information about how to process this data. As a result of differing levels of eHealth literacy (Mather and Cummings 2015), the CA must align the educational information in accordance with the user's knowledge (**MR1**) and prepare this information so that the user is able to understand it (**MR2**). Most mobile health applications provide generic health information, which may lead to misinterpretation and aggravate problems of understanding (**I2**) (Velardo et al. 2017). By including individual health information, the application can be personalised to the user's health status (**MR3**). With regard to health awareness, the following DP should consequently be considered:

DP1: *Integrate educational health information into the CA to help the user interpret the personal sensor data and health status for sustainable health awareness.*

Designing a CA differs from most common development processes because the system is intended not only to fulfil a specific purpose but also to represent a conversational and human-like experience. In this way, the interaction with a CA must appear to be natural. Adaptations of VPAs, such as Amazon Alexa and Google Assistant, have seen improvements to speech recognition and speech synthesis (Këpuska and Bohouta 2018). Nevertheless, speech comprehension remains inadequate, especially in certain domains (**I3**) (Isbister and Doyle 2002), including health awareness. Our CA must therefore be trained in understanding (**MR4**). The course of the conversation must also be natural and goal-oriented (**MR5**). Even if speech comprehension is established, interactions can have an unnatural effect if the CA has no corresponding personality (Ball and Breese 2000). The personality is represented by, inter alia, what and how something is communicated as well as the social relationship (**I4**), and the interaction is technology-focused (**I5**) (Komiak and Benbasat 2006). To avoid this, the represented personality of the CA must be aligned with the expectations of the user (**MR6**). In summary, the CA design has the following DP:

DP2: *Enable a goal-oriented and domain-specific conversation flow with a personality that meets user expectations.*

Apart from the CA interfaces, the user interfaces on other involved devices should not appear too technology-oriented, and the user requirements should be taken into account. To ensure this, the development – from the problem description to the evaluation – should be oriented towards the requirements of the user, as described, for example, in the ISO 9241-210:2010 standard (ISO, 2010). In this way, a user-centred design (**MR7**) can be achieved. Since the WHMS focuses on the vital parameters, the user interface should present these in the best possible way (**MR8**). With only the relevant information displayed, the general problem of information overload (**I6**) is reduced (van Velsen et al. 2013). Furthermore, as a consequence of increasing digitalisation, users are exposed to a large amount of data, which can overstrain them (Meister and Deiters 2015). A measure to counteract this is context awareness (**MR9**), which means that the information is not displayed on all devices at the same time, but rather made available on the preferred device depending on the situation. For example, textual information about the smartphone can be displayed if the user cannot talk, or a spoken output can be offered if the user is in a situation in which it is not practical to read. Addressing the MR regarding the user interface design, the following DP should be considered:

DP3: *Enable the CA to access various input and output devices to adapt to the user preferences and to communicate the data adequately with an information-focused interface.*

In addition to input and output devices, wearable sensors are required for a CA for WHMSs to record the vital parameters (Pantelopoulos and Bourbakis 2010). A natural language processing service is usually also required to handle the speech processing, as well as a backend server to provide the functions of data storage and processing. As a result, the entire system consists of several subsystems that run separately from one another (**I7**) (Jaimes and Sebe 2007). Apart from the influence on context awareness, the separated systems represent the reason that the individual systems must be integrated with one another (**MR10**). Doing so ensures that the individual systems interact efficiently with one another. In the special case of CAs for WHMSs, there are few standards and widespread interfaces (**I9**) (Yuehong et al. 2016), which makes development more difficult. Furthermore, given the diversity of the systems involved, which use different data structures, the data are highly heterogeneous (**I8**) (Mavrogiorgou et al. 2019). Together with the issues of the separated systems and the missing standards, the necessity of structuring and unifying the data (**MR11**) arises. With regard to the information architecture, the following DP is devised:

DP4: *Complement the CA with a system for collecting vital parameters in order to analyse the data and provide personalised information about the individual's current health status.*

To illustrate the relationships in the derivation of DPs, Figure 2 indicates which issues lead to certain MRs and how these are taken into account to derive DPs. By following the four DPs, the CA will enable all the features offered by a traditional WHMS; however, it will present the information in a more natural and conversational manner.

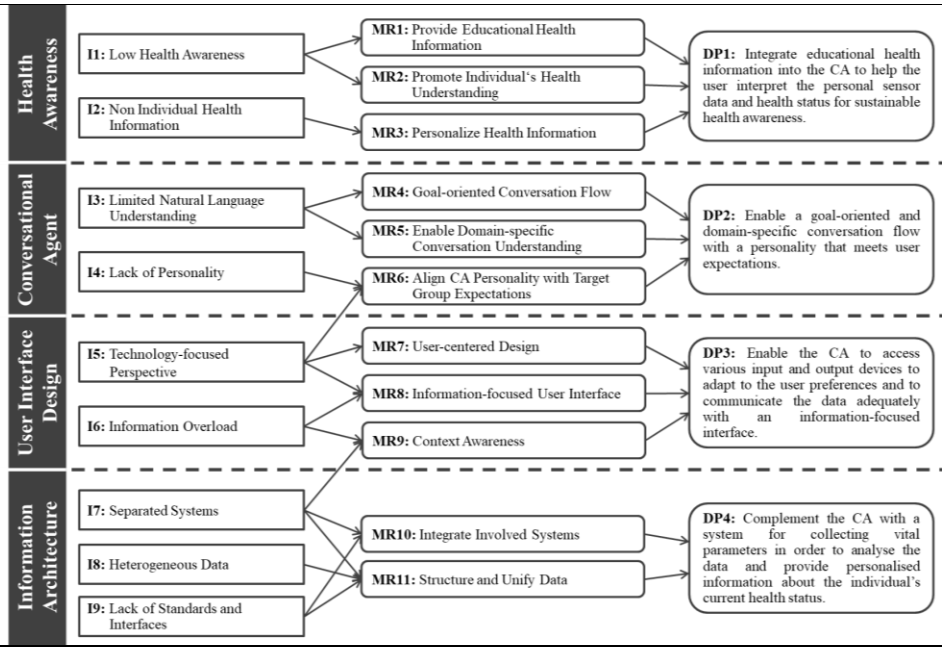


Figure 2. Issues, Meta-Requirements, and Design Principles for the CA

FeelFit Application

The DPs described in section 4 were utilised in the development of the application FeelFit. The adaptations can be divided into three areas: system architecture, CA personality, and user interface. The basis of a CA is an adaptive architecture that allows the collection of health-related sensor data (DP4) and their presentation via various input and output devices (DP3). Figure 3 illustrates the system architecture.

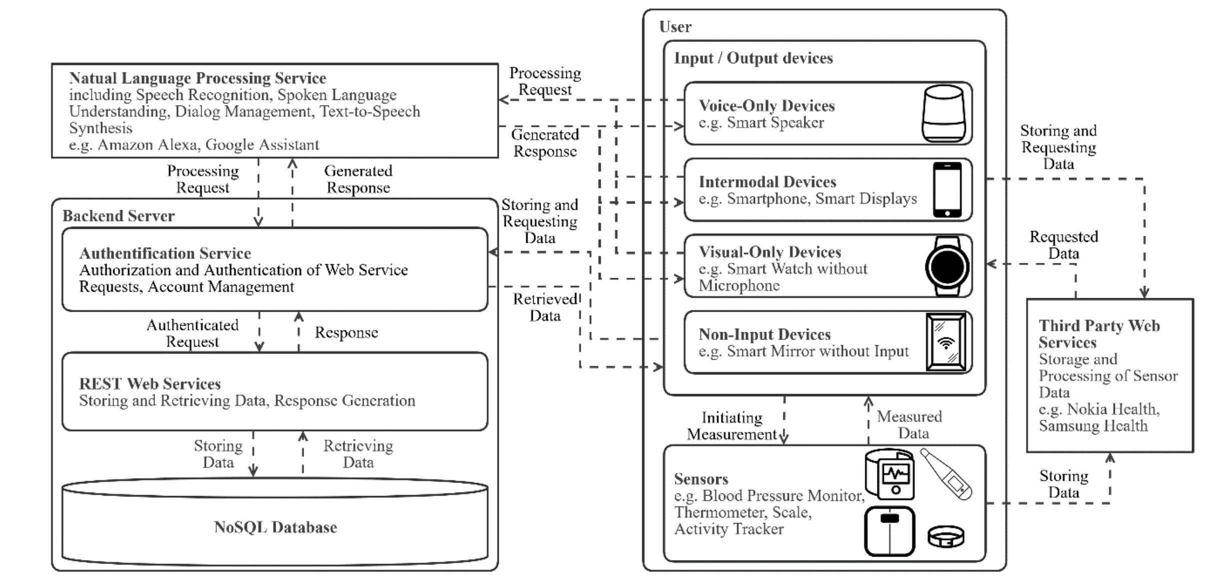


Figure 3. Architecture for a CA in a WHMS

As people can access CAs using many different devices, it is essential to categorise these devices and to analyse which will be used for the specific agent. The categorisation of input modalities by Laranjo et al. (2018) into written and spoken inputs is a valid method to classify the devices. On the one hand, written inputs can use visual-only devices, which use touch or remote input, for example, but do not allow voice input, and intermodal devices, such as smartphones and smart displays, allow inputs via voice or text. On the other hand, voice-only devices only allow voice-based inputs and do not interact with touch or written input. There are also non-input devices, which only present information to the user.

For the health monitoring system, all of these devices should be supported to provide the user with the necessary information at any time. Every device allows communication with the sensors, such as a thermometer for body temperature, a scale for weight, and a blood pressure monitor for systolic and dysostotic blood pressure readings. Apart from the direct communication between a sensor and a device via WLAN or Bluetooth, indirect communication via third-party web services must be provided as some sensor vendors only allow the use of their applications for the sensor readings. To access these sensor data, the vendors offer APIs with adequate authentication mechanisms, such as OAuth 2.0. Although it is possible for every input and output device to gather sensor data, FeelFit utilises a smartphone for the data collection as it is the most common mobile device in daily use and offers the highest processing power as well as the most functionality.

After collecting the data, it is necessary to make them available to the other devices. Therefore, REST web services on a server organise the data transfer to all devices. Since the system is personalised and involves sensitive data, an authentication service for accessing the REST services must be implemented, thereby ensuring that only previously authorised users can access their own data. All data will be stored in a NoSQL database because of its advantages over other databases in terms of storing high-frequency and heterogeneous data. Various sensor data can consequently be stored in a user's profile without making significant changes to the data structure.

Interaction with the CA using spoken or written input on voice-only, intermodal, and visual-only devices will send the input to a natural language processing service for speech recognition and (spoken) language understanding as well as dialogue management. The latter two tasks require agent-specific configuration, such as training the agent-specific classifications and conversation flows. After processing the input, a REST web service will be called upon to generate a response, and the natural language processing service will send the response back to the respective devices. For spoken input, a text-to-speech synthesis will present the response as spoken language. In addition, some inputs can trigger actions (such as display graphs on a non-input device), causing the web service to send information to another device after language processing.

The presented architecture enables a device-independent CA for a WHMS. In addition to the architecture, which illustrates the technical perspective, it is essential to create a personality for the CA that meets the expectations of the target group (DP2). As younger people are more familiar with WHMSs and CAs, the application first focuses on this group. For these users, it is generally more important to track one's vital parameters for fitness purposes than for monitoring health issues or chronic disease, and the personality of the CA must align with the goals of the users. Therefore, the CA should be designed as a fitness coach who provides the user with necessary information on his or her health status. Based on predefined and configured threshold values for the vital parameters, the CA interprets the data and offers recommendations if needed (DP1). In contrast to a CA for older people for monitoring diseases, it is not necessary to provide in-depth information with which to interpret the vital parameters, as younger users are more likely to be able to interpret the data themselves and less likely to have serious (chronic) diseases.

The personality of the CA is used to design the user interface across the devices to create an authentic experience. In essence, the CA should communicate and present the vital parameters to the user. Furthermore, the user should be informed about whether the values are normal or whether they require further examination. Since the presentation and classification of the values are in the foreground, the user interface should have a simple design that concentrates on those values (DP3). In addition, the interface should adapt to the user, since the presentation must be adapted to the monitoring of different vital parameters. The interfaces for the mobile application, CA, and smart mirror, presented in Figure 4, illustrate the implementation of an intuitive and minimalist presentation of the data. The mobile application is primarily used to connect new sensors and to set up the system. It also provides a dashboard of the configured vital sign parameters on the home screen. These can be adjusted individually and dynamically. When a vital sign is selected, the history over a specified period of time is displayed graphically. The application has been designed to ensure that all functions can be accessed from the home screen within seven clicks.

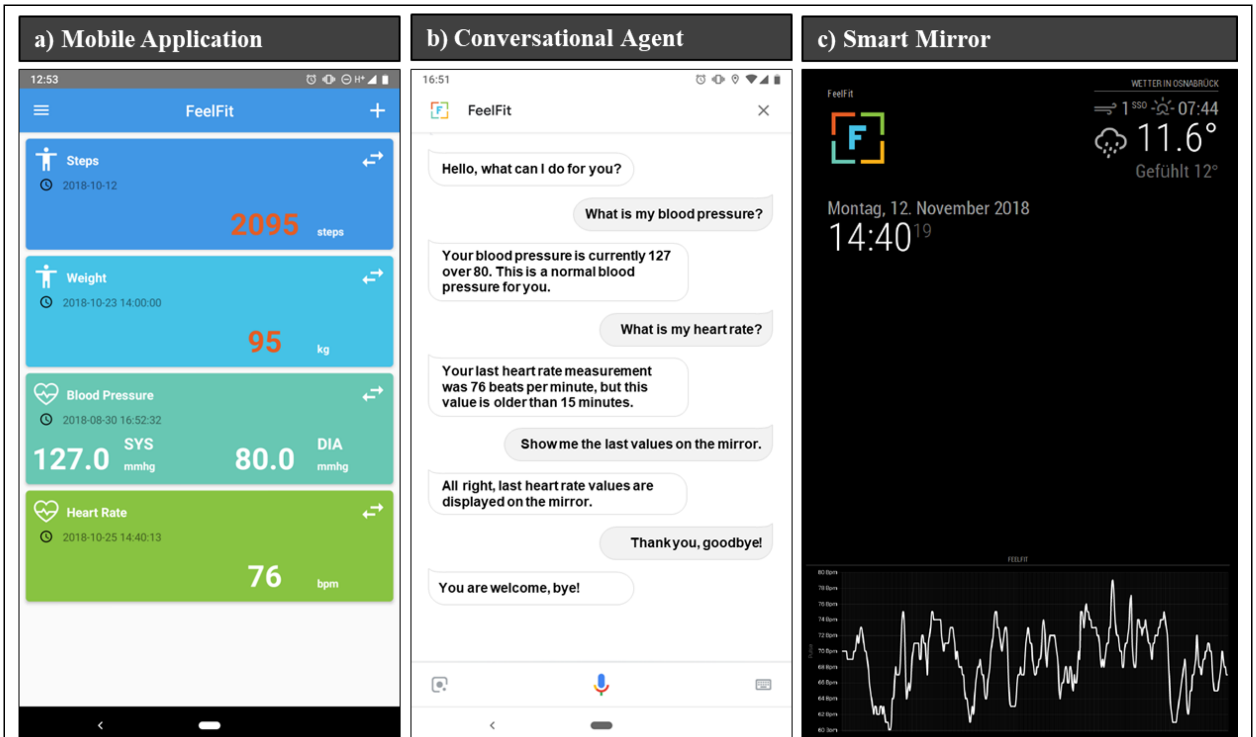


Figure 4. User Interface Design of a) Mobile Application, b) CA, and c) Smart Mirror

The CA can be used once the sensors have been set up via the mobile application. Users must initially log in with their account data, and after successful authentication they can ask the CA about their vital parameters. The CA then supplies them with the requested parameters and explains whether the data are normal for each specific user. In addition, the CA indicates whether the data are still reliable or out of date. As well as querying the vital parameters, the CA enables the data to be displayed on different devices, such as the smart mirror. The CA knows the context of the conversation and situation so that it can be used for follow-up questions, thus ensuring a more natural conversation. These interfaces are intended to provide a uniform and clear user experience according to the derived design principles. An essential component of this is the availability across all devices and in every situation. In this way, users can inform themselves about their health at any time, not only through their own subjective perception but also through vital parameters, objectively measured by sensors.

Evaluation

The integrated conversational health agent presented in the previous section is the result of three design cycle iterations. Within this process, the evaluation phases provided essential feedback regarding improvements to the application. To structure the evaluation in the different cycles, we applied a formative evaluation in three stages following the FEDS framework by Venable et al. (2016). We opted for this formative and naturalistic evaluation to validate and improve our conversational health agent in more than one evaluation step through constant feedback. Moreover, since our major design risk was user-oriented, we chose the human risk and effectiveness evaluation strategy (Venable et al. 2016). The purpose of our evaluation cycles was to address the following priorities: **usability** (**I**), **technology and CAs** (**II**), and **task fulfilment and technology acceptance** (**III**). In evaluation cycles I and III, experiments were conducted to measure the attitude of potential users, whereas in evaluation cycle II, further improvements should be identified by experts within the scope of a workshop. In each evaluation cycle, participants were introduced to the scenario and asked to perform four tasks. First, they were asked to connect a new sensor to the FeelFit application via a smartphone (1) and to take a new vital parameter measurement (2). After configuring the home screen (3), they were asked to use the CA to display data on the smart mirror (4).

In the first and the third evaluation cycles, we performed an experiment with randomly selected participants. The experiments took place in a smart home showroom within an electronic market in northwest Germany. Even though the place was equipped with Smart Home technology, it looks like a common apartment, and users only interacted with the FeelFit system. In this way, possible biases including framing, confirmation bias, and optimism bias (Fleischmann et al. 2014) are low, and the setting had two major advantages. First, we were able to avoid a laboratory atmosphere, where research experiments typically take place and where participants feel uncomfortable and observed. As the showroom was furnished as a common flat, participants felt more at ease. Second, we were able to approach customers in the market to request their participation, with these individuals assumed to have an interest in innovative electronic devices. In the experiment, we asked for personal data such as age, gender, education, and digital pre-experience (a). Participants were then introduced to the scenario in which they were asked to perform the four tasks. The second part of these experiments was a questionnaire (b). In the second evaluation cycle, we carried out a workshop with CA and technology experts in which we received feedback from the experts after the four tasks had been completed. A picture of the experiment set-up and insights into the results of evaluation cycles I and III are presented in Figure 5.

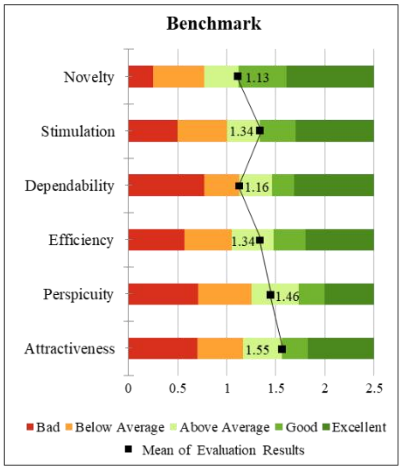
I. Evaluation cycle: usability

The first evaluation was based on the user experience questionnaire of Schrepp et al. (2014). Participants received 26-item couples (e.g., structured/unstructured, enjoyable/annoying) and used a seven-point Likert-scale to rate the technology interaction. Each item couple could thus be rated on a range from -3 to +3. Furthermore, the experimenter documented the stages at which the participants had problems completing the tasks.

a. Personal data and pre-experience

As participants were chosen randomly, there was no participant restriction. Of 54 participants, 41 were male and 13 were female. Furthermore, four of the participants were younger than 18 years old. The largest age class was 18–24 years (46.30%), and another 31.48% were aged 25–34 years. Moreover, one participant was in the class of 35–44 years, and one was between 44 and 54 years old. Finally, three participants were 55–65 years old, and three were older than 65. All participants were using either an Android-based or iOS-based smartphone, and 72.22% had previously used a CA. Digital experience was thus assumed for all participants. However, 57.41% had never used digital technologies to measure vital parameters.

Results Evaluation Cycle I



Experiment Set-up

Results Evaluation Cycle III

Task Fulfilment and Technology Acceptance

	Mean	Variance	Std. Dev.
Perceived Flexibility	1.29	1.77	1.33
Friendliness	2.04	1.52	1.23
Understandable	2.38	0.80	0.89
Comprehension	1.63	2.05	1.43
Voice Clarity	2.75	0.31	0.56
Consistency	1.42	1.87	1.37
Pleasantness	1.71	1.27	1.13
Low cognitive Demand	1.25	2.06	1.44
Consistent User Experience	1.42	1.87	1.37
Task Success	2.25	0.94	0.97
Systems Reliability	1.46	2.81	1.68
Expectation Match	1.29	1.77	1.33

Figure 5. Evaluation Design, Experiment Set-up, and Results of the Evaluation

b. User experience questionnaire

After performing the four tasks, the participants were asked to provide feedback on the CA. The responses indicated that 62.96% had a 'good' overall impression of the system, and 27.78% had a 'very good' overall impression. A further 9.26% indicated a neutral evaluation, but no respondent rated the FeelFit system negatively. The majority (72.22%) said they could imagine using the application in the future, whilst 11.11% could not. On average, all 26 items of the user experience questionnaire were rated positively. The items were consolidated into the six categories of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. The six categories achieved a mean of between 0.97 and 1.25. The overall positive evaluation of the system is supported by a small standard deviation, which is lower than 1 for each category. We achieved the highest value in the category of attractiveness, which indicates a user-friendly design. This category achieved the highest values for the requested items, such as 'good' (1.85), 'organised' (1.93), 'friendly' (1.76), and 'clear' (1.72). The lowest values were obtained for the word couples 'fast' vs. 'slow' (0.6) and 'predictable' vs. 'unpredictable' (0.85). Furthermore, the user experience questionnaire allows a benchmarking of the achieved results, with 401 different studies that used the same method with a total of 18,483 people (Schrepp et al. 2014). Compared to the benchmarking studies, our CA is above average in all categories, as illustrated on the right-hand side in Figure 5. The speed of the system was additionally criticised in the text boxes at the end of the questionnaire. In the course of further development, we have sought both to further improve our strengths and to address the weaknesses of artefacts. Among other things, we have revised the menu structure and instructions to allow the user a more intuitive use. In addition, the source code has been intensively checked and modified, and the complexity has been reduced to improve the performance of the application.

II. Evaluation cycle: technology and conversational agent

As a second evaluation cycle, a workshop with six experts from the field of technology and CAs was conducted. In addition to the revisions of the MRs and DPs presented in Figure 2, the developed applications were also discussed in-depth. Among other aspects, the experts provided helpful hints and support in improving language understanding, dialogue management, and response generation. Furthermore, options to improve the performance as well as the mobile menu design were discussed.

Apart from the technical improvement, the workshop discussed possibilities to evaluate the system with regard to the DPs. Since the first evaluation cycle focused on the usability of the system, and because the revisions of the first two cycles ensured the stable and user-friendly running of the system, the third evaluation cycle should determine whether users can inform themselves about their health status with their health-related information and become more health aware. In addition, users should compare the application with other known mobile health and fitness applications and indicate how the components of the CA and other display options improve health information delivery. To answer this question, the experts agreed that an experiment in the form of the first evaluation cycle is suitable. However, instead of usability, the focus should be on task fulfilment and technology acceptance. In addition, long-term studies could be carried out subsequently to investigate whether the CA and the associated integration into everyday life have a significant influence on health awareness.

III. Evaluation cycle: task fulfilment and technology acceptance

The third evaluation cycle focused on technology acceptance and task fulfilment in order to investigate the perceived benefit of the application. The questionnaire was based on the subjective evaluation of McTear et al. (2016). Apart from the user's background, consisting of personal, task-related, and system-related information, the subjective evaluation emphasizes on the individual interaction, the perceived task success, as well as the user's overall impression of the system. The evaluation was supplemented by questions regarding the implementation of the DPs and the added value of the individual components, such as smart mirror and CAs, in the interaction as well as a comparison with known health and fitness applications. All questions were rated on a seven-point Likert-scale from -3 to +3. Finally, the participants were asked for feedback on the improvement and extension of the application.

a. Personal data and pre-experience

Similar to the first evaluation cycle, participants were chosen randomly. The age of the participants (n = 36) ranged from 19 to 51 years, whereas gender distribution was balanced (55% male, 45% female). The largest age group was made up of 25–34-year-olds (54.76%), whereas the other participants were either younger or older. All participants were using a smartphone, and 76.19% had previously used a CA. Digital pre-experience was thus assumed for all participants. Moreover, 52.38% have already used technologies for measuring vital parameters before.

b. Task fulfilment and technology acceptance questionnaire

After the assessment of users' backgrounds, we asked the participants about their individual interaction with the system, their overall impression of the application, a comparison to other known health applications, and potential application improvements. These categories were assessed by several subcategories that are presented in Figure 5 on the right-hand side. On average, all items were rated positively, with a standard deviation that does not exceed 1.5 except for the system's reliability (1.68).

Furthermore, the respondents assessed the individual interaction as positive overall. This category is structured in several subcategories addressing primarily DP2. The respondents perceived the system to be flexible (1.29) and friendly (2.04). This is supported by the perceived clarity of the CA's voice (2.75), which received the highest rating and the lowest standard deviation (0.56) at the same time, as well as its comprehension of the user's messages (1.63) and responses delivered in an understandable way (2.38). With regard to DP3 and DP4, the respondents rated the consistency of the displayed health data (1.42), pleasantness (1.71), low cognitive demand (1.25), and a consistent user experience (1.42) in terms of the interaction with the whole system. Furthermore, the participants agreed that the system provides them with a simple information retrieval process and overview of their health status in everyday life (1.75). The effectiveness of the application is underlined by the participants' perceived task success (2.25) and system's reliability (1.46). However, as mentioned above, reliability had the highest standard deviation, which indicates that the code of the application must still be optimised. Nevertheless, the expectations were matched (1.29) by the majority of the participants. In conclusion, participants emphasised the benefit of the CA (60%), particularly that of the smart mirror (85%). At the end of the questionnaire, the respondents suggested expansions such as information sharing, nutrition tracking, different CA voice genders, and connections to other known health and fitness applications.

Results of the evaluation cycles

After conducting each evaluation cycle, we derived potential improvements from the received feedback. A summary of identified issues is listed in Table 1. As expected in evaluation cycles I and III, older participants faced problems performing the tasks, since the majority of these participants had no prior experience with a CA or wearables. However, with little guidance, all participants were ultimately able to complete the tasks. No participants aged less than 45 years had problems completing all four tasks. During the scenario, we noticed that some participants who suffer from a chronic disease, felt reminded of their hospital treatment by looking at the visualisation of their vital parameters on the smart mirror. As a solution, the interface of the smart mirror might be revised to appear less clinical. In addition, as a prerequisite for further evaluation in a field test, the system's reliability must be improved. To address as many users as possible, the variety of the available sensors must be increased. In this way, users could participate in a long-term evaluation with the sensors they need for their individual goals and then provide feedback on the impact the application has on their individual health awareness that is not based on a short experiment.

Table 1. Feedback of the formative evaluation

Evaluation cycles			Feedback
I	II	III	
■			Improve feedback after completing a step (e.g., coupling with the sensor)
■			Align colour design
■			Add user-specific critical limits when displaying vital parameters
■			Avoid application disruption
■	■		Align structure of the main menu
■	■		Simplify layout
■	■		Increase performance or speed of application
	■		Provide the CA with more functions
	■	■	Improve system reliability to enable long-term usage evaluation in a field study
	■	■	Improve flexibility for communication with the CA
		■	Provide option to choose the type and gender of the CA's voice
		■	Implement the sharing of data with healthcare professionals
		■	Provide health-related information (e.g., meaning of specific vital parameter)
		■	Implement daily reminder for medicals
		■	Implement connection to other health and fitness apps (such as nutrition tracking)
		■	Enhance sensor variety
		■	Match same parameter data of different sensors
		■	Analyse data for deeper insights

Discussion

In each iteration of this study, the experts of the workshops, as well as the participants in the experiments, positively evaluated the idea of using a CA to monitor vital parameters and its implementation. All the experts and a substantial majority of the respondents said that, in particular, the natural language interaction and the integration of different sensors in a central instance had high added value. For example, an older participant noted that he saw '*clear advantages for less IT-savvy people*' because they do not have to navigate through multiple applications.

In addition, we asked both the experts and the participants in the evaluation study for critical feedback on the DPs. With regard to DP1, one participant said that '*many users do not have an overview of their vital parameters*' and consequently have little understanding of the importance of those parameters and how they can be influenced (positively). Another participant pointed out that a high level of trust could be placed in such an application and that this could lead to a '*false sense of security*', which could then lead to health damage. We counteract this, however, by advising users to consult medical specialists, if necessary, in the case of exceeded thresholds. We also state, at the launch of the application, that this is not a medically approved product and that it primarily serves the purpose of health monitoring. During the workshop in the second iteration, one expert remarked that he often had a problem with VPAs; he did not receive the information he expected, and in some cases, follow-up-questions were not possible or not recognised in context. As a consequence, DP2 is useful, improving the user experience by answering requests in an effective and goal-oriented manner. This was emphasised by the users, who, in the course of the evaluation, rated the understanding of the CA and the conduct of conversations to be positive. Furthermore, in the user experience questionnaire, the aspects of perspicuity, efficiency, and dependability were evaluated positively and indicate a high pragmatic quality (Schrepp et al. 2014). Therefore, by following DP2, future CAs can achieve a high pragmatic quality.

In contrast, users have different experiences with CAs. Older people, in particular, deal with CAs differently, compared to so-called digital natives, for example. The challenge of the need to take appropriate (IT) knowledge into account when designing a CA thus arises. Concerning DP3, one participant stressed that, for her, '*the inhibition threshold was significantly lowered*' because both the input and output devices could be selected based on one's preference, and a central instance was established in which the data were aggregated and processed. In principle, however, this freedom also makes it possible for incorrect data to enter the system through the devices. During a workshop in the first iteration, an expert described the integration of heterogeneous systems as a challenge, whilst another remarked that he found it '*complex and strenuous*' to open various applications to obtain an overview of his vital parameters, which he then had to evaluate himself. Based partially on this criticism of existing applications, we derived DP4. Participants in the study praised the implementation of DP4 as '*superfluous clicks through various applications*' were no longer necessary, and the voice-based operation was described as '*instinctively usable*'.

Overall, our findings indicate that CA can be a useful alternative or supplement to conventional graphical user interfaces, especially for users with limited computer experience. Table 2 summarises the key findings of our evaluation with end users and the discussion with experts in the last workshop.

Table 2. Key findings regarding the DPs

DP	Key Findings
DP1	▪ Enables users to react independently to changes in their vital parameters (such as high blood pressure) ▪ Ensures that the CA is not perceived to be a healthcare professional – a healthcare professional should be consulted for further information if certain parameters have conspicuous values
DP2	▪ Encourages users to interact more with the CA (e.g., by asking follow-up questions) ▪ Ensures that (a) users focus on the main purpose of the application and (b) the target group is addressed
DP3	▪ Establishes a manageable, personalised hub by aggregating and editing the collected health data ▪ Lowers the inhibition threshold to use the CA by free selection of input and output devices ▪ Enhances users' understanding by providing targeted, comprehensible answers ▪ Intercepts incorrect inputs by the devices
DP4	▪ Presents the user with personal information on their current health status in natural language ▪ Reduces the effort required to manoeuvre manually through applications by synthesising the available sensor data and additional information

Through the development, implementation, and evaluation of the DPs, the presented DPs contribute to the information systems discipline by providing important guidance in designing successful CAs for practical challenges, for example health awareness in everyday life (Diederich et al. 2019). The multiple refinements of the DPs in each iteration as well as the design and evaluation of the instantiation FeelFit ensures the validity of the DPs. This study aligns with studies by Gnewuch et al., who have defined DPs for CA applications for customer feedback (Gnewuch et al. 2017) and energy feedback (Gnewuch et al. 2018). The application-oriented DPs supplement the general DPs, which focus the context-specific design of CA user interaction (Nunamaker et al. 2011) The CA thus supports domain-specific processes in the most effective way. With our research, we also contribute to the use of CAs in the field of personalised healthcare, especially telehealth, by providing individual vital parameter readings and providing necessary information about vital parameters and individual thresholds. Furthermore, the modular and device-independent architecture allows for a convenient integration into everyday life and can be utilised as a basic architecture for further CAs. The identified MRs and DPs also offer a starting point for further research and can be critically examined by researchers and tested in more realistic application scenarios. In addition, they can be extended to contribute to the scientifically founded development of CAs. Moreover, this paper contributes to work on the basic design of CAs by presenting our application.

Apart from the contribution to the knowledge base of CA design and users' health awareness, this paper has highly practical relevance as the developed DPs offer added value in the design of CAs. The developments in this area (e.g., Amazon Alexa skills, distribution of smart home hardware) demonstrate a need for scientifically based design criteria. In addition, the transparent development process and presented architecture (Figure 3) illustrate the possibilities for companies that wish to engage in this domain. The evaluation results clearly demonstrate that the use of natural language-based CAs represents added value for users (at least in this context) and therefore support the need for more entrepreneurial commitment in this area.

Alongside the advantages outlined above, the developed artefact is associated with challenges and risks. First, from a data protection perspective, all recorded data are stored centrally in one place. Not only does the centrality of data storage represent a risk in terms of data loss, but it can also offer an attack surface for targeted data manipulation. Furthermore, fundamental challenges arise from a (mis)interpretation of the results by the user, since he or she might not have adequate eHealth literacy to interpret the measured data. Too narrow or too generous thresholds could lead to serious problems for patients as well as healthcare system. One ethical challenge identified is potential dependence on such applications. In addition, the extent to which the CA reduces 'real' social contacts (for example with healthcare providers, such as doctors) and the effects this has on users would need to be investigated. Legal issues must also be addressed in this context. The question of who could be held responsible for any health-related consequential damage arises with regard to the (incorrect) reading, processing, output, and interpretation of vital parameters. The application developed in this research project should be seen primarily as a supplementary application, without the claim to analyse medical findings, but rather to inform on the current state of health and fitness.

Furthermore, VPAs remain in an early stage of development. In addition, it can usually not be ensured that the person using the application via a VPA is also the person logged in. Therefore, it is necessary to add further security measures on the part of VPA providers, so that the system knows whom it is speaking to and which information may be outputted. Nevertheless, the evaluation with test subjects demonstrates that the use of CAs in this area offers many advantages, especially in connection with visual representations.

Conclusion

We developed an integrated CA in a one-year research project. This work was motivated by recognition of the fact that many people do not track their individual health status (Palumbo 2017), whilst the popularity of wearables to measure different vital parameters rises, and VPAs appear more in everyday life. Our system is intended to (a) integrate different sensors to monitor vital parameters, (b) offer the possibility to query (c) output data by a CA, and (d) display the data on various devices.

To answer the RQ, we analysed the literature and existing solutions on the market and identified nine relevant issues and 11 MRs, from which we derived four DPs. Based on the DPs, we developed the application FeelFit, which integrates various sensors to measure vital parameters, stores data in a database, and makes the data available on a smartphone, a CA, and a smart mirror. So far, the system is able to measure heart rate, blood pressure, temperature, weight, and breathing frequency. In three evaluation iterations, we investigated the usability and task fulfilment of the CA with a total of 90 experiment participants as well as the CA design in a workshop with six experts, and we discussed the implications of using a WHMS and a CA.

In addition to increasing health awareness, the system has the potential to support the management of chronic diseases and to adjust medications with sensors of medical accuracy. Patients can easily share their data with healthcare professionals; a longitudinal observation of vital parameters can allow physicians to improve diagnoses and monitor the wellbeing of their patients. At the same time, increased health awareness might be useful to increase patient adherence. The evaluation confirmed the novelty and the usability of our solution, which indicates that the CA fulfils our derived DPs.

Limitations and Future Work

Despite the promising results of our study, our presented solution has limitations. First, using a CA with different sensors cannot replace a professional clinical diagnosis. The measurement of vital parameters should rather be seen as a supplement to tracking one's health status. Nevertheless, the measured data might help physicians and pharmacists to improve their diagnoses, with the collected data integrated into electronic health records and shared with healthcare professionals. Physicians would be able to continuously track critical vital parameters and define individual thresholds. A second limitation is that the utility of the CA depends on external sensors. Without interfaces from sensor producers, the system could not provide major benefits. Furthermore, the accuracy of today's sensors is not sufficient for medical use. However, the open architecture design allows for the integration of additional sensors, so that even more vital parameters can be gathered.

Overall, our CA represents a useful solution for increasing personal health awareness. The system offers numerous possibilities for extensions, such as additional sensors, a connection to electronic health records, and integration of nutrition tracking. Moreover, artificial intelligence methods can be applied to analyse vital parameters and enable predictions about changes in a user's health status. In addition to improving the system, future research opportunities lie in the analysis of ethical, legal, and social implications of using a CA to monitor individual health status. Since the evaluation revealed the willingness of at least young and technologically adept people to use a WHMS with a CA in everyday life, we will focus our future work on improving the system's reliability so that long-term studies can examine the impact of the system on users' health awareness. Finally, future studies should investigate how systems such as FeelFit can be implemented in the healthcare system to realise the identified potentials of enhanced patients' health awareness.

Acknowledgements

The authors would like to thank the experiment participants, the project team Pia Schüttler, Bastian Kamp, Nils Kohl, Simon Lübbers and Shaowei Yu, for their valuable and substantive help, as well as the reviewers for their constructive feedback. This contribution was prepared within the research projects "Dorfgemeinschaft 2.0" (German Federal Ministry for Education and Research, funding code 16SV7453) and "Apotheke 2.0" (German Federal Ministry of Food and Agriculture, „Land.Digital", PT BLE).

References

ADA. 2019. "Ada: Your Personal Health Guide." (https://ada.com/, accessed April 17, 2019).

Ball, G., and Breese, J. 2000. "Emotion and Personality in a Conversational Agent," *Embodied Conversational Agents*, pp. 189–219.

Baumgart, R., and Wiewiorra, L. 2016. "The Role of Self-Control in Self-Tracking," in *Proceedings of the 37th International Conference on Information Systems*, Dublin.

Bickmore, T. W., Utami, D., Matsuyama, R., and Paasche-Orlow, M. K. 2016. "Improving Access to Online Health Information with Conversational Agents: A Randomized Controlled Experiment," *Journal of Medical Internet Research* (18:1), JMIR Publications Inc., p. e1.

Blaschke, T. F., Osterberg, L., Vrijens, B., and Urquhart, J. 2012. "Adherence to Medications: Insights Arising from Studies on the Unreliable Link between Prescribed and Actual Drug Dosing Histories," *Annual Review of Pharmacology and Toxicology* (52), Annual Reviews, pp. 275–301.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., Brocke, J. Von, and Reimer, K. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," in *Proceedings of the 17th European Conference on Information Systems 2009*, Verona, pp. 2206–2217.

Dey, A. K., Abowd, G. D., and Salber, D. 2001. "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications," *Human–Computer Interaction* (16:2–4), Taylor & Francis, pp. 97–166.

Diederich, S., Brendel, A., and M Kolbe, L. 2019. "On Conversational Agents in Information Systems Research: Analyzing the Past to Guide Future Work," in *Proceedings of the 14th International Conference on Wirtschaftsinformatik*, Siegen, February 24, pp. 1550–1564.

DOCYET. 2019. "DOCYET." (https://docyet.com/, accessed April 17, 2019).

Fitzpatrick, K. K., Darcy, A., and Vierhile, M. 2017. "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 Mental Health* (4:2), JMIR Publications Inc., p. e19.

Fleischmann, M., Amirpur, M., Benlian, A., and Hess, T. 2014. "Cognitive Biases in Information Systems Research: A Scientometric Analysis," in *Proceedings of the 22nd European Conference on Information Systems*, Tel Aviv.

Giorgino, T., Azzini, I., Rognoni, C., Quaglini, S., Stefanelli, M., Gretter, R., and Falavigna, D. 2005. "Automated Spoken Dialogue System for Hypertensive Patient Home Management," *International Journal of Medical Informatics* (74:2–4), Elsevier, pp. 159–167.

Gnewuch, U., Morana, S., Heckmann, C., and Maedche, A. 2018. "Designing Conversational Agents for Energy Feedback," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Cham: Springer, pp. 18–33.

Gnewuch, U., Morana, S., and Maedche, A. 2017. "Towards Designing Cooperative and Social Conversational Agents for Customer Service," in *Proceedings of the 38th International Conference on Information Systems (ICIS)*, Seoul, South Korea.

Griol, D., Carbó, J., and Molina, J. M. 2013. "An Automatic Dialog Simulation Technique to Develop and Evaluate Interactive Conversational Agents," *Applied Artificial Intelligence* (27:9), Taylor & Francis, pp. 759–780.

Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quaterly* (28:1), pp. 75–105.

Hudlicka, E. 2013. "Virtual Training and Coaching of Health Behavior: Example from Mindfulness Meditation Training," *Patient Education and Counseling* (92:2), Elsevier, pp. 160–166.

Isbister, K., and Doyle, P. 2002. "Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy," in *The First International Joint Conference on Autonomous Agents & Multi-Agent Systems*.

Jaimes, A., and Sebe, N. 2007. "Multimodal Human–computer Interaction: A Survey," *Computer Vision and Image Understanding* (108:1–2), Elsevier, pp. 116–134.

Jiang, R. 2019. "Introducing New Alexa Healthcare Skills." (https://developer.amazon.com/de/blogs/alexa/post/ff33dbc7-6cf5-4db8-b203-99144a251a21/introducing-new-alexa-healthcare-skills, accessed April 29, 2019).

Këpuska, V., and Bohouta, G. 2018. "Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 99–103.

Kim, H., and Xie, B. 2017. "Health Literacy in the EHealth Era: A Systematic Review of the Literature," *Patient Education and Counseling* (100:6), Elsevier, pp. 1073–1082.

Klopfenstein, L. C., Delpriori, S., Malatini, S., and Bogliolo, A. 2017. "The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms," in *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17*, ACM, pp. 555–565.

Komiak, S. Y. X., and Benbasat, I. 2006. "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," *MIS Quarterly*, JSTOR, pp. 941–960.

Krippendorff, K. 2004. "Reliability in Content Analysis," *Human Communication Research* (30:3), Wiley Online Library, pp. 411–433.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., and Coiera, E. 2018. "Conversational Agents in Healthcare: A Systematic Review," *Journal of the American Medical Informatics Association* (25:9), pp. 1248–1258.

Luger, E., and Sellen, A. 2016. "Like Having a Really Bad PA: The Gulf between User Expectation and Experience of Conversational Agents," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, pp. 5286–5297.

Lupton, D. 2016. *The Quantified Self*, John Wiley & Sons.

Maass, W., Kowatsch, T., Janzen, S., and Varshney, U. 2011. "A Natural Language Technology-Enhanced Mobile Sales Assistant for in-Store Shopping Situations," in *Proceedings of the 19th Europeen Conference on Information Systems*, Helsinki: AIS Association for Information Systems.

Mather, C., and Cummings, E. 2015. "Empowering Learners: Using a Triad Model to Promote EHealth Literacy and Transform Learning at Point of Care," *Knowledge Management & E-Learning* (7:4), The University of Hong Kong-Faculty of Education, p. 629.

Mattila, E., Pärkkä, J., Hermersdorf, M., Kaasinen, J., Vainio, J., Samposalo, K., Merilahti, J., Kolari, J., Kulju, M., and Lappalainen, R. 2008. "Mobile Diary for Wellness Management—results on Usage and Usability in Two User Studies," *IEEE Transactions on Information Technology in Biomedicine* (12:4), IEEE, pp. 501–512.

Mavrogiorgou, A., Kiourtis, A., Perakis, K., Miltiadou, D., Pitsios, S., and Kyriazis, D. 2019. "Analyzing Data and Data Sources towards a Unified Approach for Ensuring End-to-End Data and Data Sources Quality in Healthcare 4.0," *Computer Methods and Programs in Biomedicine*, Elsevier, p. 104967.

Mayring, P. 2010. "Qualitative Inhaltsanalyse," in *Handbuch Qualitative Forschung in Der Psychologie*, Springer, pp. 601–613.

McTear, M., Callejas, Z., and Griol, D. 2016. *The Conversational Interface: Talking to Smart Devices*, Springer.

Meister, S., and Deiters, W. 2015. "Information Logistics Solutions to Cope with Big Data Challenges in AAL and Telemedicine," in *Ambient Assisted Living*, Cham: Springer, pp. 77–88.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., and Patton, M. W. 2011. "Embodied Conversational Agent-Based Kiosk for Automated Interviewing," *Journal of Management Information Systems* (28:1), Taylor & Francis, pp. 17–48.

Palumbo, R. 2017. "Examining the Impacts of Health Literacy on Healthcare Costs. An Evidence Synthesis," *Health Services Management Research* (30:4), SAGE Publications Sage UK: London, England, pp. 197–212.

Pantelopoulos, A., and Bourbakis, N. G. 2010. "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* (40:1), IEEE, pp. 1–12.

Paré, G., Poba-Nzaou, P., and Sicotte, C. 2013. "Home Telemonitoring for Chronic Disease Management: An Economic Assessment," *International Journal of Technology Assessment in Health Care* (29:2), Cambridge University Press, pp. 155–161.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), Taylor & Francis, pp. 45–77.

Plattner, H., Meinel, C., and Weinberg, U. 2009. *Design Thinking*, (H. Plattner, C. Meinel, and L. Leifer, eds.), Landsberg am Lech: Springer.

Prugger, C., Keil, U., Wellmann, J., De Bacquer, D., de Backer, G., Ambrosio, G. B., Reiner, Z., Gaita, D., Wood, D., and Kotseva, K. 2011. "Blood Pressure Control and Knowledge of Target Blood Pressure in Coronary Patients across Europe: Results from the EUROASPIRE III Survey," *Journal of Hypertension* (29:8), LWW, pp. 1641–1648.

Rhee, H., Allen, J., Mammen, J., and Swift, M. 2014. "Mobile Phone-Based Asthma Self-Management Aid for Adolescents (MASMAA): A Feasibility Study," *Patient Preference and Adherence* (8), Dove Press, pp. 63–72.

Schrepp, M., Hinderks, A., and Thomaschewski, J. 2014. "Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios," in *International Conference of Design, User Experience, and Usability*, Cham: Springer, pp. 383–392.

Seeger, A.-M., Pfeiffer, J., and Heinzl, A. 2018. "Designing Anthropomorphic Conversational Agents: Development and Empirical Evaluation of a Design Framework," in *Proceedings of the 39th International Conference on Information Systems*, San Francisco.

Stanford University School of Medicine. 2019. "Results Of A Large-Scale, App-Based Study To Identify Atrial Fibrillation Using A Smartwatch: The Apple Heart Study," in *American College of Cardiology (ACC)*, Late-Breaking Clinical Trial 1, New Orleans.

Swan, M. 2009. "Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking," *International Journal of Environmental Research and Public Health* (6:2), Molecular Diversity Preservation International, pp. 492–525.

Tanaka, H., Negoro, H., Iwasaka, H., and Nakamura, S. 2017. "Embodied Conversational Agents for Multimodal Automated Social Skills Training in People with Autism Spectrum Disorders," *PloS One* (12:8), Public Library of Science, p. e0182151.

Velardo, C., Shah, S. A., Gibson, O., Clifford, G., Heneghan, C., Rutter, H., Farmer, A., and Tarassenko, L. 2017. "Digital Health System for Personalised COPD Long-Term Management," *BMC Medical Informatics and Decision Making* (17:1), BioMed Central, p. 19.

van Velsen, L., Beaujean, D. J. M. A., and van Gemert-Pijnen, J. E. W. C. 2013. "Why Mobile Health App Overload Drives Us Crazy, and How to Restore the Sanity," *BMC Medical Informatics and Decision Making* (13:1), BioMed Central, p. 23.

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: A Framework for Evaluation in Design Science Research," *European Journal of Information Systems* (25:1), Taylor & Francis, pp. 77–89.

Watson, A. R. 2016. "Impact of the Digital Age on Transforming Healthcare," in *Healthcare Information Management Systems*, K. J. Weaver C., Ball M., Kim G. (ed.), Cham: Springer, pp. 219–233.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:1), XIII–XXIII.

Yuehong, Y. I. N., Zeng, Y., Chen, X., and Fan, Y. 2016. "The Internet of Things in Healthcare: An Overview," *Journal of Industrial Information Integration* (1), Elsevier, pp. 3–13.