# Communicability of traditional interfaces VS chatbots in healthcare and smart home domains

**Article** *in* Behaviour & Information Technology · June 2019

DOI: 10.1080/0144929X.2019.1637025



**3 authors**, including:

Stefano Valtolina
University of Milan



Barbara Rita Barricelli
University of Brescia

Behaviour & Information Technology

ISSN: 0144-929X (Print) 1362-3001 (Online) Journal homepage: https://www.tandfonline.com/loi/tbit20

# Communicability of traditional interfaces VS chatbots in healthcare and smart home domains

Stefano Valtolina, Barbara Rita Barricelli & Serena Di Gaetano

**To cite this article:** Stefano Valtolina, Barbara Rita Barricelli & Serena Di Gaetano (2019): Communicability of traditional interfaces VS chatbots in healthcare and smart home domains, Behaviour & Information Technology, DOI: 10.1080/0144929X.2019.1637025

**To link to this article:** https://doi.org/10.1080/0144929X.2019.1637025

Published online: 30 Jun 2019.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2019.1637025

# Communicability of traditional interfaces VS chatbots in healthcare and smart home domains

Stefano Valtolina, Barbara Rita Barricelli and Serena Di Gaetano

[a]Department of Computer Science, Università degli Studi di Milano Milano, Italy

**ABSTRACT**
This paper presents a study about communicability of conversational interfaces (namely chatbots) under a semiotic perspective. A chatbot is a software system that allows you to simulate real conversations between devices and users by means of a conversational interface (CI). After introducing the chatbot concept, focusing on its advantages and issues, we will present two domains of use in which chatbot interfaces can be effective: healthcare and smart home. For carrying out simple tasks such as finding information or triggering operations, users need an easy-to-use and to an easy-to-learn system to communicate with. To face this, conversational interfaces represent the latest trend in the field of digital design. For studying the communicability aspects of a CI, we carried out a user test to compare traditional and chatbot interfaces. This paper aims at evaluating the benefits at the communicability level of a chatbot in comparison to traditional GUI for incrementing the effectiveness and efficacy of communication between users and the system specifically for users with poor attitude in using technologies. In details, we evaluated the communicability of two prototypes that can be used to solve simple tasks in order to favour user inclusion, including everyone with very little exposure to technologies.

**ARTICLE HISTORY**
Received 16 November 2018
Accepted 22 June 2019

**KEYWORDS**
Conversational interfaces; design trade-off; smart home; healthcare; communicability; computer semiotics

## 1. Introduction

In the current era of connectivity and technological innovation, user experience (UX) is becoming a more and more complex problem to deal with in the development of interactive systems. Problems such as the digital divide and social exclusion can only serve to exacerbate difficulties in accessing interactive systems. The emphasis on technology excellence without much attention to greater social inclusion and user engagement resulted only in minor improvements in participation.

While economical effort and other product-related constraints will often dictate which interface style to use for a given application, the interaction type that will best support UX can highlight potential trade-offs, dilemmas, and pros and cons. A huge part of UX is the interface that sits between a user and a service. In an attempt to make the human–computer interaction more efficient, designers continuously try different approaches. In the past, one of the most effective solutions was to provide users with a GUI (Graphical user interface) populated by colourful icons, menus, lists and windows in order to make information accessible, understandable and usable. Nevertheless, with the advent of the era of Internet of Things (IoT), Ambient Intelligence (AmI), Big data, Quantified Self, new models are needed that can deal with a lot of data, unstructured information, text, and services. From this perspective, traditional GUIs have one major problem. They are artificial creations invented to enable interactions between human and computers. Users have to adapt to interfaces for learning rules on how to operate with them and sometimes they are overwhelmed by information and navigation structures that move the user's attention away from the task at hand.

These limitations refer mainly to issues related to social inclusion and digital divide that lead to a failure in handling technology or to a poor attitude in using digital systems for solving also simple problems. Sometime ICT-based systems may raise barriers and create inequalities between digitally-included and digitally-excluded users who find it difficult to use (Barricelli et al. 2016; Dirks, Bühler, and Edler 2018). In this field, the challenge is to deliver personalised, tailored, and frictionless experiences when people need them most. For this reason, it is important to push boundaries and think about new ways to use technology.

To face this problem, the conversational interface (Zue and Glass 2000) is the latest trend in the field of digital design that is focused on improving how people interact with systems by means of a more natural communication way. Industry leaders such as Apple, Google, Microsoft, Amazon, and Facebook are strongly focused

**CONTACT** Stefano Valtolina valtolin@di.unimi.it Department of Computer Science, Università degli Studi di Milano, Via Celoria, 18, 20133 Milano, Italy

on building a new generation of systems designed around human conversation strategies. The more an interface leverages human conversation, fewer users have to be taught how to use it (Følstad and Brandtzæg 2017). Conversational interfaces are designed to allow users to speak or to chat with bots by means of voice (voicebot) or text (chatbot) (Shawar and Atwell 2007). The human-system dialogue consists of a user looking for information, and the agent (bot) providing it. The main difference between a chatbot and a voicebot is the way users can interact with them. A chatbot provides users with a text-based dialogue like the one typically used on messaging platforms, including SMS, social network systems and web-based applications. This means users interact with chatbots on a screen by using rich user interfaces endowed with buttons, menus or other graphic items. On the other hand, users interact with a voicebot using their voice, i.e. in natural language. The voicebot then answers back using pre-recorded messages, text-to-speech responses or a combination of both (Braun et al. 2017). In some cases, speaking is the most efficient way to communicate especially in situations where a person cannot use their hands (e.g. in case of voice assistants for drivers). The efficiency of these agents is based on dialogue features that include nuance and tone, emphasis and pacing. These type of agents has to embed the typical person's voice, on which the conversational tone is modulated. Chatbots are less dependent on the need to provide a bot with 'human abilities' such as a personalised tone of voice or a speech fuelled by emotion and they are more focused on providing a chat-based conversation. Nevertheless, nowadays differences between chatbots and voicebots are becoming increasingly blurred. Some conversational interfaces provide a screen-based interaction as input strategy or in alternative users can dictate using their phone's text-to-speech feature. In this way, the bot may be available as a skill integrated into a voice-activated chatbot.

Regardless of the interaction style used by the bot, today several domains can benefit from this type of virtual agent: e-commerce, health-care assistants, customer service systems or IoT device management. By exploiting the ubiquity of mobile devices, messenger bot can provide a lower barrier of entry to the user since today texting is one of the most dominant modes of communication. Users interact with a screen for selecting an option in a list (e.g. for choosing a product to buy or an exercise to carry out) or for acting on a device (e.g. for regulating the light level of the lamps in the living room or the thermostat temperature).

Although chatbots are widely used, their diffusion is recent and related literature is still limited. In particular, existing literature focuses mainly on the description of conversational agents and the related pros and cons. We noticed a lack of literature about the effects of conversational agents in terms of usability and UX provided by this kind of interaction.

Given our involvement in two Projects named 'I-VITAE New Pathways for Life' and 'Social appliances for Industry 4.0 – EaSy 4.0 – Smart Living', we had the possibility to study the benefits of using conversational agents in two specific domains: Smart home and Healthcare. These are two typical fields in which chatbots can provide a more effective communication strategy specifically for users with poor attitude in using technologies. Motivating users to live healthier lives or helping them to control home devices are realistic examples of challenges that today conversational agents have to deal with. In this field, our research questions are: Are chatbots able to provide a communication strategy that better fits the user's mood and needs in comparison to traditional mobile-based applications?

On the basis of our research experience (Zhu et al. 2010; Ardito et al. 2011; Zhu, Barricelli, and Iacob 2011; Barricelli and Valtolina 2017), we believe an effective communication strategy might help all users in focusing on solving their problems rather than spending time in trying to understand complicated interfaces or interaction styles. In particular, for users who are not well-accustomed to technology use, but who need support for changing bad attitudes or behaviours (e.g. quit smoking or follow a diet) or interested in receiving recommendations or suggestions (e.g. for buying a train ticket or cooking a meal), digital agents can help to improve their quality of life. It is just in such situations that we want to study the positive effects of the communication mediated by chatbots.

According to these considerations, how to evaluate the benefit at the communicability level of a chatbot in comparison to traditional GUI? In other words, which criteria can we use for testing the way the bot triggers behaviour change, such as engagement, considering user needs and burdens?

For measuring the communicability of a system, we adopted a Communicability Evaluation Method (CEM), which is based on the semiotic engineering theory (De Souza 2005). Computer semiotics is the science of signs that are at the basis of designing interface signs. The study of semiotic aspects is required to construct well-designed user interfaces and achieve the desired quality of communicability, which in turn improves usability. In this field, we are interested in studying how chatbots are able to communicate with the user in a more effective and efficient way with respect to the communication strategies carried out by traditional interfaces. In summary, the objective of our research is to explore the communicability of systems based on chatbots and traditional GUIs. For all we know, this is the first study that aims to compare these two types of interaction styles, in particular through the investigation of their communicability aspects by means of Semiotic Engineering methods.

### 1.1. Structure of this paper

The rest of this paper is structured as follows. Section 2 investigates related works about advantages and obstacles in using chatbots for making the communication effective. In Section 3, we outline the requirements, problems, and potentials in adopting chatbots in two contexts of use such as Smart home and Healthcare. Moreover, we describe two prototypes we developed that are based on conversational user flow diagrams, which respectively 1) help women in monitoring menstrual cycle or 2) support users in setting and controlling their washing machine(s). Section 4 describes our experiments aimed at investigating the UX according to usability, user satisfaction and the effectiveness of the communication between users and bots. Section 5 describes the results of our evaluations and section 6 presents a discussion about the main findings and limitations of our study. Finally, Section 7 concludes this paper and shares our vision regarding the benefit of using a chatbot for providing a better communication strategy between users and systems.

## 2. Related works

'Before there were computers, we could distinguish persons from non-persons on the basis of an ability to participate in conversations. But now, we have hybrids operating between person and non-persons with whom we can talk in ordinary language' (Colby 1999).

Chatbots can be used for a wide range of fields, such as education (Letzter 2016), information retrieval (Shawar, Atwell, and Roberts 2005), business and e-commerce (Chai et al. 2001), and for customer service (MarutiTechlabs 2017). In their study, Tatai et al. (2003) compared implementations of chatbots and identified three main roles of chatbots, namely: (i) Digital Assistant; (ii) Information Provider; (iii) General Chatbot (Tatai et al. 2003). The development of this technology is what transformed traditional interfaces into conversational interfaces. Different types of chatbots can be identified, based on their learning capability, the way in which they interact with the user, and the context of the application. Most advanced chatbots are able to perform actions based on the conversations (for example, chatbot for reserving a table). The technicality behind the bot architecture could be simple or sophisticated. For example, a chatbot system could use machine learning, neural networks, wearables or IoT devices to handle tasks, or it could be rule-based and use a finite state machine to accomplish very simple tasks. Intention-based agents understand language as commands, and they use that understanding to perform tasks. Well-known examples of intention-based agents include Amazon's Alexa, Google Home, Apple's Siri and Microsoft's Cortana. Understanding what the user says as a command requires solving two problems: (i) Identifying what the user wants the machine to do (the 'intent'). (ii) Figuring out the details of the intent so the machine can take action. The assistant can determine the intent using either keywords or text-based classification. To use keywords, it is necessary to simply associate words and phrases with intents. To do text-based classification, it is necessary to label a set of statements with the correct intents and then train a classifier over them. Conversational agents expand on intention-based agents to have multi-turn conversations. To do this, they must keep track of the state of the conversation and know when a person wants to talk about something else (Venkatesh et al. 2018).

### 2.1. Advantages and strength of chatbots

Chatbot success is due to different dynamics related both to the relationship between companies and customers and to the development of technology and smartphones/mobile devices.

Regarding the relationship between companies and customers, the reason that led to the spread of chatbots is given by a change in the communication between companies and users. Companies indeed are adapting their channels of communication, mainly their websites, to different devices, first of all, mobile devices. Websites are often considered difficult to manage by the companies and to visit by the users, so many companies developed mobile apps in order to exploit the potentials of new tools. Today, the mobile apps' market is saturated and the competition is high. Strategies implemented to face competition are mainly focused on providing easy and intuitive interfaces. Companies' goal is to foster loyalty to address the problem of market volatility by involving consumers directly. Within this paradigm, chatbots are tools that allow companies to scale mobile messaging with users, facilitating the conversation with them. For this reason, the use of chatbots is growing and it is foreseen that between 2017 and 2023 the spread of chatbot will increase up to 37% (Chatbots Market Research Report- Global Forecast 2023). One of the main advantages for the users is given by the ease of use of chatbots because they work as an instant messaging application.

As previously mentioned, the second reason that ensured the success of chatbot is the development of technology and the spread of smartphones. Smartphones are today used for several needs and consulted many times during the day. The instant messaging mechanism is familiar to most of those who own smartphones, thanks to the huge success of apps, such as Whatsapp, Telegram and Facebook Messenger. Through the chatbot interface, companies can create easy and pleasant conversations, leading to a more positive and engaging UX (Nocera et al. 2015; Brandtzaeg and Følstad 2017). Compared to traditional channels of communication, such as telephone calls and emails, chatbots are faster and always available, ensuring quick and easy answers related to different problems. Given the possibility to manage unlimited data and users, chatbots can also replace assistance service operators, who are not always available. As a recent report by My Clever Agency demonstrates (M. C. Agency, Chatbots: a consumer research study, 2016), chatbot is preferred by 84.6% of interviewed people as consumer channel to get answers, second only to face-to-face interactions, but still preferred to email or online form, chat with operators, phone, and social media.

Another positive implication of chatbots is the possibility to simplify the information search process. Today people have access to a big amount of information, due to the increase of devices and channels of communication. Studies demonstrate that the abundance of information generates an increase in distraction: too many sources of information lead people to share their attention to different services, reducing the average time dedicated to every single source (Torchiani 2018). One of the main goals for companies is to catch the attention of consumers and users, optimising the short time available. In this sense, chatbots allow companies to delete redundant information, creating interfaces where consumer/user asks in natural language and receives a concise and appropriate answer (Shawar and Atwell 2007). In addition to the simplification of the conversation, chatbots provide also the possibility to personalise the conversation with users, exploiting personal information obtained by previous interactions (James 2016). Thanks to Sentiment analysis tools, it is possible to adapt the conversation to the mood or the attitude of the users (Jongeling, Datta, and Serebrenik 2015). This aspect will be further discussed in Section 3.1.3.

Finally, a further positive aspect from companies' point of view is that implementation of basic chatbots is neither expensive nor time-consuming: there are several open source platforms that give companies the possibility to create and personalise a chatbot in a short time.

### 2.2. Obstacles and barriers to the diffusion of chatbots

Besides the above described positive implications, it is necessary to mention the main obstacles to the diffusion of chatbots. The main barrier is related to the nature of conversational agent intelligence (Robino 2018). Bots are designed to follow a specific path and they rarely accommodate deviations away from a programmed script. If the conversation with the user becomes too complex, a chatbot is likely to have problems in recognising and understanding users' requests and consequently, it does not provide the right answers. In order to avoid this issue, the ability of artificial intelligence software needs to be improved through Natural Language Understanding and Machine Learning algorithms. Another possible obstacle to take into consideration is the reluctance to the use of chatbot by users (Jenkins et al. 2007). The resistance of change is a common dynamic that occurs when a new technology is introduced. Furthermore, in some specific cases, users could be reluctant to use chatbots if they are asked to provide information that is particularly sensitive, for example, personal information or bank account details. Chatbots may expose the users to unwelcome push notifications and spam. Finally, it is necessary for companies to consider chatbots as conversational platforms: as in a real conversation, it is fundamental not to interrupt the talk and avoid distractions that may also be annoying for the users.

### 2.3. Chatbots: how they can make the communication effective

In order to better understand what is the added value for the communicability in using conversational interfaces, we need to investigate how human dialogues work. Taking into account the pros and cons of using a chatbot our study aims at understanding if a conversational agent is able to provide and support better communication between users and systems. Specifically, if it is able to engage users with poor attitudes in using technologies.

The work in (Valtolina, Barricelli, and Dittrich 2012) outlines the computer-semiotics terms that can be used to identify problems in developing knowledge-management systems to support collaboration across heterogeneous domains. This work highlights how digital communication processes derive from the work of Ladislav Tondl (Tondl 1981) on analogue communication processes. Tondl's model was adapted and expanded in (Marcante and Mussio 2006) to be applied to digital communication and to clarify the role of human actors in the human-machine communication process. Under this perspective, we can describe the digital communication process from a semiotic point of view as an individual who is sending a message to a second individual, the system, which tries to respond to the user's request by the mediation of the. This model is framed in a computer-semiotics context and can be used to analyse the communication process of any digital interfaces. According to Andersen's definition (Andersen 1997), computer semiotics is 'a discipline that analyses computer systems and their context of use under a specific perspective, namely as signs that users interpret to mean something'. From this perspective, Semiotic Engineering (SE) (De Souza et al. 2006) views interactive software systems as artefacts through which the communication between users and systems takes place. The system sends users 'a one-shot message' – which unfolds into further two-way message exchanges, according to de Souza (De Souza et al. 2006) – explaining how and why they should communicate with the software application in order to achieve a specific goal.

According to this point of view, if the interface is designed to structure navigation and contents in ways unfamiliar to the user, the user will not be able to use the software for the task at hand or to communicate with other users in a collaboration context. As Norman stated in (Norman 1992), 'The real problem with the interface is that it is an interface'. In other words, the users use products to solve specific problems, and when they solve a problem, they want to focus on the problem itself, not in understanding the interface. This is particularly true when users, especially the ones who are not accustomed to using technology, have to carry out simple tasks such as to find basic information or to trigger simple commands. From a SE point of view is the interface itself that makes the process of problem-solving harder because it introduces a cognitive load on top of the problems. The most comfortable and useful interfaces are the closest to natural human communication, and the most human natural interface is the spoken language.

According to this consideration, the contribution of our study aims at investigating how conversational interfaces are able to introduce an opportunity to interact with a machine in a more natural way. To evaluate this hypothesis, we need to study conversational interfaces according to the effect they have on the communication between the users and the bot.

Results from some studies have highlighted how conventional usability evaluation methods like heuristic evaluation (Nielsen and Mack 1994), and even exploratory methods like the cognitive walkthrough (Cathleen et al. 1994), do not reflect the opinions of the users (Thompson and Kemp 2009) (Gomes da Silva and Dix 2007). For example, in studies run by Gomes da Silva and Dix in 2007, it was found that YouTube failed when tested using heuristic evaluation although it is one of the most popular Web apps. Moreover, another study presents a thesis according to which, in some cases, focusing on usability can be harmful (Greenberg and Buxton 2008). This is because usability methods tend to put the lens on the usability bugs and not on the whole usefulness of an application. Innovative ideas could be discouraged by negative results, which lead to giving up on plans that might otherwise bear good fruits. For this reason, our idea is to evaluate the conversational interface by applying a method of semiotic engineering evaluation: the communicability evaluation method (CEM) (Prates, de Souza, and Barbosa 2000) (De Souza and Leitão 2009). This method can be used for going beyond the limit of cognitive engineering methods. CEM method explores the communication, trying to identify through users' observation the empirical evidence of the effects that the system's messages have on the users' interaction. The chatbot designer's objective is to let users reach their goals by interacting with the interface, and thus, in the SE perspective, achieve effective user-system communication. Therefore, the communication is effective when the user interacts with the system according to the designer's project, using the strategies, understanding what the system can do and how. Obviously, this communication is not always effective: It is the case of a communicability breakdown, an interruption of user-system communication. Communicability breakdowns are errors, misunderstandings, interaction failure: every time the user has some problems with interacting with the system.

## 3. Chatbots in two domains of use

For the purpose of our analysis, we envisage two domains of use in which chatbot can be effective: a) Healthcare and b) Smart home. We studied current mobile apps belonging to these domains, underlying main features and, reasons why these apps could exploit chatbot technology to be more efficient. The aim of our analysis is not to give a classification of healthcare and smart home apps available today, but rather to underline requirements and positive implications they might have.

### *3.1. Healthcare*

During the last years, many healthcare mobile apps spread in several fields, from the monitoring of physical activity to the reservation of a doctor's appointment. The success of these apps is due to two different orders of reasons. First of all, the spread of awareness about the importance of being healthy and maintaining healthy habits led to the need for monitoring people lifestyle and putting more attention to everyday routine. Health Promotion can be described as 'the process of enabling people to increase control over, and to improve, their health" (W.H.O. 1995). Given the general attention on lifestyle promotion, people are more interested in monitoring regular activities, such as meals, exercise, hours of sleep. In the same way, the increase of elderly people led to the need for healthcare support even after their treatment period. The second order of reasons is related to the development of new technologies that allow facilitating the process of the state of health monitoring. The usage of smartphones and the introduction of a wide variety of sensors are enabling the spread of personal health apps that can actively monitor, model and promote wellbeing.

The collection of data about health could be useful for caregivers because it provides doctors with different information about their patients, avoiding caregivers to spend a vast amount of their time documenting their patients' condition. Healthcare apps allow t reducing the caregivers' workload regarding data collection, reminders, user follow-up, engaging users in pre-built scripts and predictable conversations. In the context of the project named 'I-VITAE New Pathways for Life' we had the possibility to study the use of mobile apps for the monitoring of the menstrual cycle. Most of these apps give users the possibility to track their periods, ovulation, contraception, and reproduction. Furthermore, they are designed as a personal diary where users can record specific information such as cervical mucus, BMI, sexual activity, weight, temperature, symptoms or mood. Having a collection of this information could be helpful for building a higher self-awareness or, more practically, in the event of gynaecological problems, because it is possible to provide specialists with different data that can help them identify or exclude certain diseases.

#### 3.1.1. Open issues

For the purpose of this study, we do not focus on the difficulty to develop medical knowledge-based technologies. The aim is rather to point out critical aspects related to the interaction with healthcare mobile apps. From the interaction point of view, the main critical aspects to take into consideration concern the sensitivity of the issues related to personal state of health and the necessity of acting on users' behaviours. Healthcare mobile apps help users deal with issues that sometimes are hard to face, such as infertility or weight problems. If users are facing a difficult situation, sometimes talking about their problem could be frustrating or painful and they are not motivated to use the app. In this sense, it is necessary to take into consideration a possible resistance by the users to disclose their personal information. Regarding the second issue, the goal of healthcare apps is not limited to the moment of usage of the app but it extends to users' daily routine. The long-term goal is to engage users in healthy activities to promote their lifestyle, leading to a behaviour change (Grasso, Cawsey, and Jones 2000).

#### 3.1.2. Requirements

Literature review shows that the capacity to motivate people to use the application is one of the most important requirements for an application that faces the issue of lifestyle promotion. Motivation to guarantee is double: On the one side, users have to be encouraged to put into practice all those behaviours considered healthy (diet, physical activity, meditation, etc.); on the other side, they have to be motivated to collect and report their results and feedback. Indeed, the usefulness of this kind of apps is given by the possibility to create a user's profile and complete tracking of their data. For facing sensitive data and collecting different kinds of information, a possible solution is to engage users by leveraging the simplicity of the technology, integrating it with users' daily routine. The effectiveness of the simple approach is well studied in the literature (Fogg 2009) (Oinas-Kukkonen and Harjumaa 2018). The goal is to guarantee simplicity of interaction between user and system from the first usage. Once the user has acquired necessary skills, using the system becomes easier, and this further enhances their ability because the more familiar behaviour is, the more likely the user is to do it (Fogg 2009). Furthermore, users might have some difficulties in comprehending the meaning of some data related to health because they might not have the required level of medical knowledge. For this reason, it is necessary to simplify also the information given and requested by the app, for example providing explanations and additional information. Because of the need to develop applications that are engaging and at the same time provide information, most of the current health apps have a specific common feature: the presence of a coach. Through a coaching approach, it is possible to deliver tasks and information and to remind users to collect and register data. Interaction is simplified because the virtual coach is seen by the users as a doctor or an expert to turn to.

#### 3.1.3. Chatbots for healthcare

For the reasons described above, we investigated the healthcare field as an application domain to develop chatbot apps. Given the simplicity of interaction ensured by a chatbot, a conversational agent could be a valid tool to track individual health activities. The possibility to receive and send messages is familiar to most of the users and reduce the level of complexity linked to the activity of monitoring data. Conversational agents aim to facilitate the interaction with users by means of conversation and request/response interaction. The aim is not to substitute the caregiver, but to adopt a coaching approach and to support doctors' activity. A chatbot acts as a virtual coach interacting with the users, reminding them to do their tasks and to report their results. Through the dialogue, it is possible to optimise user personalisation and natural interaction. According to a study (Boratto et al. 2017), users tend to be more engaged to train when their training is developed and remotely supervised by a coach. Another factor to take into consideration is the previously mentioned sensitivity of the issues related to healthcare. In this sense, a chatbot could be a useful interface solution because it allows one to create a friendly conversation that facilitates to face sensitive issues. The solution provided by the chatbot is given by the possibility of personalising the tone of voice and to adapt the conversation to the user's mood, through the Sentiment Analysis. Sentiment analysis is a method to identify the attitude of participants in an online discussion. An attitude is the mental position of one participant with regard to another participant, to a product or to the discussed topic, and it could be positive or negative. Sentiment analysis is used in a wide range of domains, from the evaluation of product reviews on the Web (Morinaga et al. 2002; Turney and Littman 2003) to the analysis of political speeches. Through this analysis, it is possible to predict whether a sentence displays a positive or negative attitude and to adapt the dialogue with the user depending on their attitude. When sentiment analysis reveals a negative users' attitude towards a product, such as negative reviews, the company has the possibility to change the product itself or the marketing strategy. In the context of healthcare applications, if the analysis displays a negative mood, such as reluctance to reveal personal information or problems, chatbots might change the tone of the conversation, making users feel comfortable. The conversational agent is abstract enough to cover a vast number of domains in the context of lifestyle promotion and it is possible to design several hypotheses of dialogue, addressing the dialogue towards physical activity, diet promotion or mental wellness. While mental wellbeing is usually evaluated with questions on the person's emotional, affective and cognitive status (e.g. 'How are you feeling now?'), physical activity is assessed through objective and factual information (e.g. 'How many steps did you take today?'). For this reason, another requirement to develop a successful chatbot is to integrate emotional and factual information. Integration of different behavioural and motivational factors of an individual into the chatbot knowledge base is a key factor to generate support tailored to individual's needs and preferences in a specific situation (Lindgren, Guerrero, and Janols 2017). Healthcare chatbots currently might be a mix of both patient-only (apps that help a patient track and make sense of health data) and patient-caregivers applications (apps that connect the two groups, for diagnosis and/or treatment).

For the purpose of our analysis, given our involvement in a project based on this topic, we decided to evaluate the communicability level of apps developed within the field of menstrual cycle monitoring.

After conducting an analysis about existing mobile apps that help users in tracking their menstrual cycle, we chose to test 'Il mio calendario'[1], one of the most spread apps in Italy developed by Abishkking. It is free, available both for Android and for iOS and it has been positively evaluated by users (4.9/5). 'Il mio calendario' provides users with the possibility to insert information about their menstrual cycle: starting date, duration, symptoms, etc., creating a sort of medical records of the user. Starting from the traditional interface of this app, we developed a chatbot with the same features provided by the traditional one. In order to be easily tested, we call it 'Il mio calendario' as well. It is based on a conversation and request/response interaction, in order to help users in tracking the information that could be useful to collect. The chatbot asks for specific information (for example basal temperature, last period, cycle length, etc.) giving users the possibility to skip some questions and to reply at a later time. Furthermore, based on users' answers, the chatbot deepens certain topics and dynamics. For example, if the user states to be tired, the chatbot investigates the reasons for this symptom asking questions about her quality of sleep. In this way, chatbot 'Il mio calendario' allows creating a friendly conversation that facilitates to face sensitive issues. Furthermore, it integrates emotional and factual information: in addition to specific factual data related to the menstrual cycle (for example the cycle length), it collects information related to the emotional state of users and mood.

### 3.2. Smart home

Smart home refers to the process through which appliances and devices available in a house can be remotely controlled by users. Devices are interconnected through the internet and, in order to control the devices, users need to use a mobile or another networked device and to be in any internet-connected place in the world. Centralised control of lighting, heating, alarms, air conditioning, home entertainment, houseplant, and yard watering are examples of the smart home. The spread of smart home is due to different reasons. One of them is the development of different technologies belonging to the paradigm of the Internet of Things (IoT). For the purpose of this paper, we do not deepen the concept of IoT, but we consider necessary to underline that the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction is one of the first steps that leads to the concept of the smart home. Different smart systems have been proposed where the control is e.g. via Bluetooth (Yan and Shi 2013), internet (Tan, Lee, and Soh 2002), short message service (SMS) (Khiyal, Khan, and Shehzadi 2009). Another positive implication of smart homes could be related to elderly and disabled people because they might be helped in controlling devices and appliances despite their physical impairments. In other words, a smart home can increase the quality of life for persons who might require the help of caregivers, increasing their level of independence (Chourabi et al. 2012).

Starting from the results of a project named 'Social appliances for Industry 4.0 – EaSy 4.0 – Smart Living' in which we are involved, we studied the problems that users have to control the automation of washing machines. In this field, an IoT app can help users in better washing laundry by finding the right balance between high-quality washings, save money and energy consumption.

#### 3.2.1. Open issues

Security risks and bugs are the main challenges faced by smart home technology. Expert hackers, for example, can gain access to a smart home's internet-enabled appliances. A possible solution to reduce the risk of hacking is to adopt strong passwords to protect smart appliances and devices, using encryption when available and only connecting trusted devices (Wilson 2018). Another obstacle to the diffusion of smart homes is their perceived complexity; if users have difficulty with the technology they likely give up on it with the first annoyance. In this sense, it is necessary to reduce complexity and to improve UX to encourage users with different levels of technological knowledge.

#### 3.2.2. Requirements

Despite the wide-scale efforts to popularise IoT, it still offers many practical challenges. IoT dashboards are often saturated with various data, charts, and tables making it difficult for users to find the required information. A key challenge concerns the possibility to manage data generated by IoT sensors in an efficient and easy way. The main problem is that data generated by the sensors are large but also diverse (varying in quality and type) and multimodal (temperature, light, sound, video, etc.) in nature. If the management of IoT data is one challenge, drawing insights from the data and being able to present it in a timely, understandable way is a much larger challenge. It is at the data analysis step that the true value of any IoT application is determined, and this is where Artificial Intelligence (AI) plays a crucial role by making sense of data streamed from devices (Dirican 2015). AI serves to detect patterns in this data from which it can learn to adjust the behaviour of the IoT service. Traditional methods of analysing structured data are not designed to efficiently process the vast amounts of real-time data that stream from IoT devices. This is where AI-based analysis and response becomes critical for extracting optimal value from that data. The problem of managing and analysing IoT data can be solved through many solutions and cloud platforms, which offer storage and computing infrastructure to accomplish the task. Existing IoT cloud solutions handle data sources and transmission challenges. However, a major challenge of existing IoT systems is conveying data about the various devices back to the user in a simple humanly understandable way. This requires communication strategies that allow to understand the true intent of the user query and to process information from the environment. Moreover, these solutions have to access a global network of information via the internet and have to be easily programmed to retrieve information in real-time.

#### 3.2.3. Chatbots for smart home

In order to design successful smart home apps, it is necessary to make them simple to use and intuitive. Given the difficulty to approach smart home devices, related mobile apps should be a means to simplify the interaction. Another important requirement to be met is a high level of security. As previously mentioned, vulnerabilities have been discovered in personal smart home devices and in the protocols that operate between those devices. Current mobile apps belonging to the smart home paradigm are used to control a specific set of devices produced by a company.

For the purpose of this paper, we studied possible mechanisms through which users can easily communicate with their appliances. In this sense, as partners of the 'Smart Living' Project financed by Regione Lombardia, we based our analysis on the web application developed in the context of this Project. The main goal of the application is to help and guide people in using their appliances. At the basis of the Project, there is an artificial intelligence algorithm that provides specific settings when a user needs to set their appliances such as the washing machine or the oven. Thanks to the algorithm, users can receive the best combination of parameters to set their appliances based on opinions and feedbacks given by other users. Users are asked to insert information about their necessity. For example, if they want to do the laundry, they have to indicate the typology of cloth they want to wash, the level of dirt, the quantity of clothing, etc. In the same way, if they need to set the oven, they have to indicate what they are going to cook, the quantity of food, etc. When users use the appliances according to the settings suggested by the application, they are asked to provide feedback on the results. Also, in this case, we developed a chatbot version of the traditional application. In particular, we focused only on the usage of the application to set a washing machine. The chatbot creates a dialogue to ask user information about the laundry (colours of the clothes, material, level of dirt …) and it proposes a set of parameters to set. If the user accepts the suggestion, the bot reminds the user to give feedback about the wash.

## 4. User study

In the previous Section, we outlined some solutions based on the use of conversational interfaces for supporting users in accessing health information or for controlling devices. In this field, our contribution aims at understanding how the conversation agents are able to provide users with a communication strategy that better fits their needs and expectations.

For the experimental part of our analysis, we performed a user test on mobile apps for healthcare and smart home domains. For each domain, we selected two apps, one having a traditional interface and the other using a chatbot. For the user study, we applied a combination of semiotic and cognitive evaluation methods for measuring communicability, usability and user experience (UX).

### *4.1. Materials*

#### *4.1.1. Healthcare*

For the healthcare domain, we used an existing app with a traditional interface and we designed and prototyped a chatbot-based app. The app with the traditional interface is ‘Il mio calendario’ (see Figure 1), one of the most used apps for menstrual cycle tracking in Italy. The main features that the app offers are: Tracking of the duration of the menstrual cycle, logging of symptoms, push notifications about the arrival of the menstrual cycle and fertile days. Whether users have irregular or regular periods, they may consider useful to have a specific calendar where to insert such information.

‘Il mio calendario’ becomes a sort of personal diary but also a medical record that can provide useful information related to users’ health. In fact, the app allows the user to track the assumption of contraceptive pills, basal temperature, symptoms, mood, the entity of the menstrual flow and sexual intercourses. The application collects all the data within a Calendar (depicted in Figure 2) that shows menstrual cycle days and that foresees the different level of fertility during the month and the next menstrual cycle starting day.

On the basis of the experience developed in using ‘Il mio calendario’ app, we developed a mobile application that provides the same features of the traditional one but is based on a conversational interface (chatbot), and we called it ‘LovApp’. We built ‘LovApp’ by using Google Assistant technology through the Actions on the Google developer platform. Actions on Google is a developer platform that has been letting us to integrate into our solution the functionality of the Google Assistant, Google’s virtual personal assistant. Specifically, we used Actions on Google to easily create and manage delightful and effective conversational experiences between users and our app. From a technical point of view, we started with the definition of a ‘persona’ to associate to our bot. A person who can assist users in what they can do and that can help them in understanding how the bot works on the base of what users already know. In our case, the persona was modelled as a female assistant —trustworthy with women’ needs and personal information. The metaphor of the assistant makes this new experience feel familiar. Once we defined in a clear way who is communicating and what they are communicating about, we created the dialogue. Dialogs are the key to creating great Actions on Google and they convey the flow that the user will actually experience. By writing sample dialogs, we experimented with and evaluate different design strategies, such as how to gather personal information or how to confirm a user’s request (e.g. for asking to provide the data about the menstrual cycle). For designing the flows of dialogue, we started by sketching them using the Draw.io flowchart tool.[2]

Once the dialogs were defined, we implemented the building blocks at the basis of the ‘LovApp’ agent. In details, we specified: (i) The actions, that is the entry points of any user’s interaction; (ii) The intents, that is, the underlying goals the user want to achieve or the tasks she can do (e.g. registering her basal temperature or providing data about her menstrual cycle). In Actions on Google, these are represented as a unique identifier and the corresponding user utterances that can trigger the intent; (iii) The Fulfillments, that is,

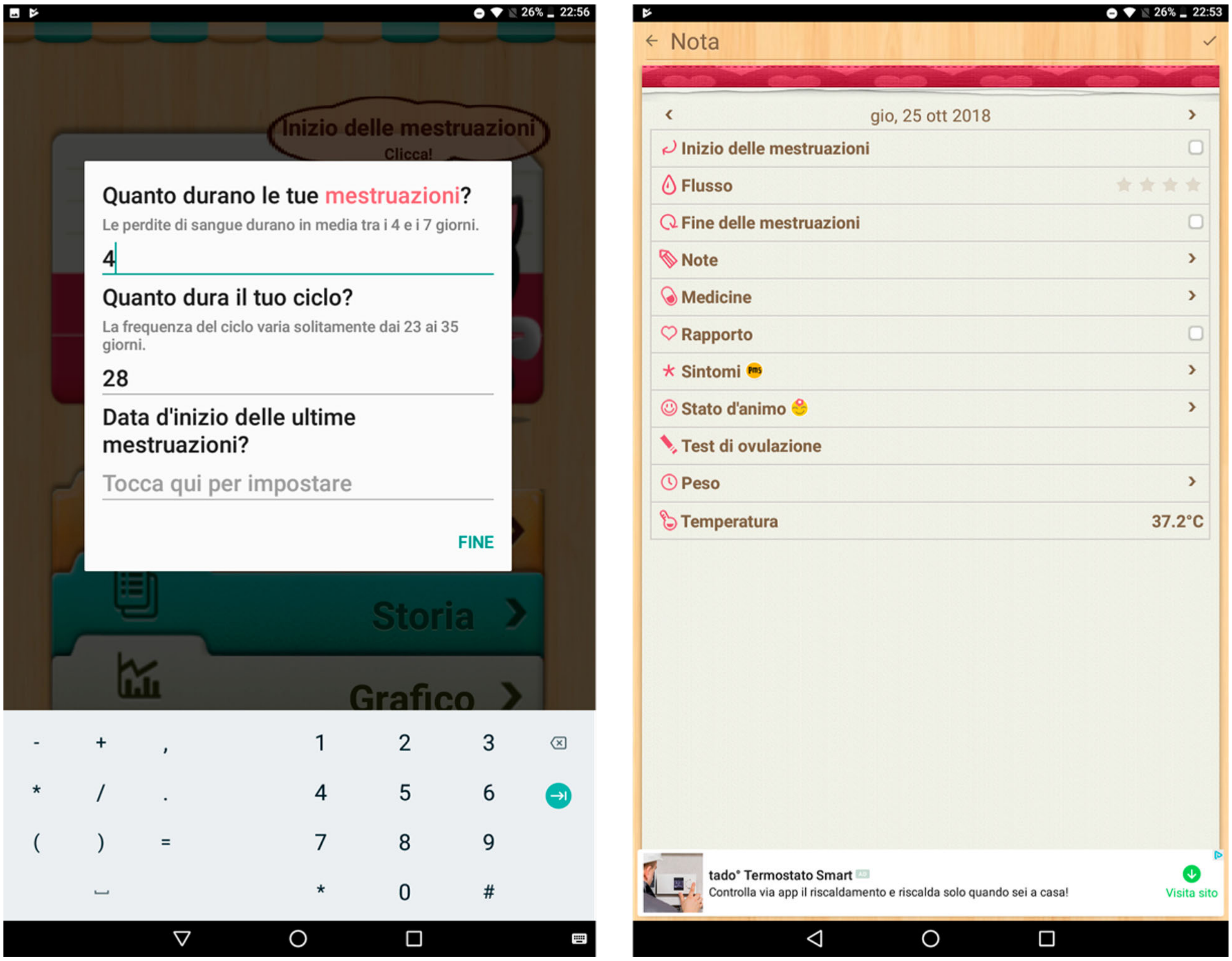


**Figure 1.** Two screens of 'Il mio calendario' app. On the left screen, the user is asked to respond to questions about the duration of the period, days between two periods and last period date. On the right screen, the list of other information that the user can provide.

the JavaScript functions that we developed for handling the intents and return an intelligent or useful response. Fulfilment is custom logic that we implemented as a webhook, which requests services, processes them, and returns responses. The conversational flow we designed, after the login, as shown in Figure 3, allows the agent to start the dialogue by asking questions to the user in order to create her profile (e.g. menstrual cycle duration, last period date).

After collecting this preliminary information, the chatbot presents more specific questions (Figure 4), like symptoms (for example headache, fever, muscle pain, and chills), mood, and basal temperature. When the user states to have a negative mood or symptom, such as to be tired or sleepy, the chatbot deepens the reason for this mood providing further questions. For example, if the user states to be tired, the application asks her whether she slept well. At the end of the questions, the user can choose to insert further details or to visualise her calendar.

### 4.1.2. Smart home

For the smart home domain, we used two apps designed and developed in the frame of Project EaSy 4.0, which serve the same purpose but differ in terms of interfaces: The former has a traditional interface ('SmartLiving'), while the latter is chatbot-based ('SmartLiving Chatbot'). Both apps support the user in setting and controlling washing machines, optimising costs and energy consumption. The apps are not linked to a specific brand or device but provide information and suggestions useful to optimise the usage of any product. The suggestions are the result of an artificial intelligence algorithm that collects users' feedbacks and returns the best washer cycle.

The typical usage scenario of 'SmartLiving' consists in a sequence of screens in which the users are asked to provide information about their washing machine (one or more, if they have more than one residences) and other information like the number of people living in the same place. Brand and model are the main


**Figure 2.** Menstrual cycle calendar in 'Il mio calendario' app.

information required by the app to find out the settings that the appliance can offer to the users, and how the machine can be set up for specific washer cycles.

Another important information that the app uses for building its suggestions is the user's most important concern:

- Time: the user prefers quick washer cycles
- Energy: the user prefers to reduce the energy consumption
- Cost: the user prefers to cut the costs related to the usage of the machine (not only energy but also soap and other additives)
- Quality: the user is more interested in the quality of the washing than in reducing time, energy or costs.

When the user starts to load the washing machine and needs advice, the 'SmartLiving' app asks to provide information about the type of clothes that are to be washed (Figure 5). Such information is related to the composition of the textiles (e.g. cotton, wool, denim), how dirt are the clothes, how coloured they are, and how loaded is the machine.

Once all this information is given, the 'SmartLiving' app provides a suggestion: Information about the washer cycle to set, spin speed, and temperature. The user can decide whether to accept or not the suggestion (see Figure 6). If the user accepts it, he/she is asked to provide, at the end of the washer cycle, feedback that evaluates:

- cleaning level;
- the humidity of the clothes;
- soap residues;
- the colour maintenance;
- softness;
- clothes shrinkage.

Since the artificial intelligence algorithm that was developed in the Project EaSy 4.0 is the basis of both the apps we developed, the 'SmartLiving Chatbot' app asks for the same information used by the 'SmartLiving' app, but letting the user interact with a conversational interface. In this case, we implemented the chatbot by following the same design strategy presented for the 'LovApp' agent previously described. We started with the definition of a 'persona' to associate to our bot. In this case, the persona was modelled as a humanised washing machine for portraying an image of a laundry expert. Then we created the sketches of the flows of dialogs by using Draw.io and finally, we implemented the actions, intents, and fulfilments that compose the building blocks at the basis of the conversational interface of our bot.

Initially, the 'SmartLiving Chatbot' app asks the user brand and model of their washing machine. All other information, like the most important criterion for washer cycle selection and the characteristics of the clothes loaded in the machine, are asked sequentially (as depicted in Figure 7).

After this, the app provides a washer cycle suggestion and, if the user accepts it, a message indicating the estimated wash time appears. The app gives the user the possibility to choose whether to insert feedback about the last suggestion (Figure 8) or ask for another suggestion.

### *4.2. Participants*

We involved two groups of participants of 20 users each that evaluated the apps of the two domains distinctly. The 40 users were invited to participate and we presented them the information sheet and explained the protocol and motivation of our study. Before starting with the

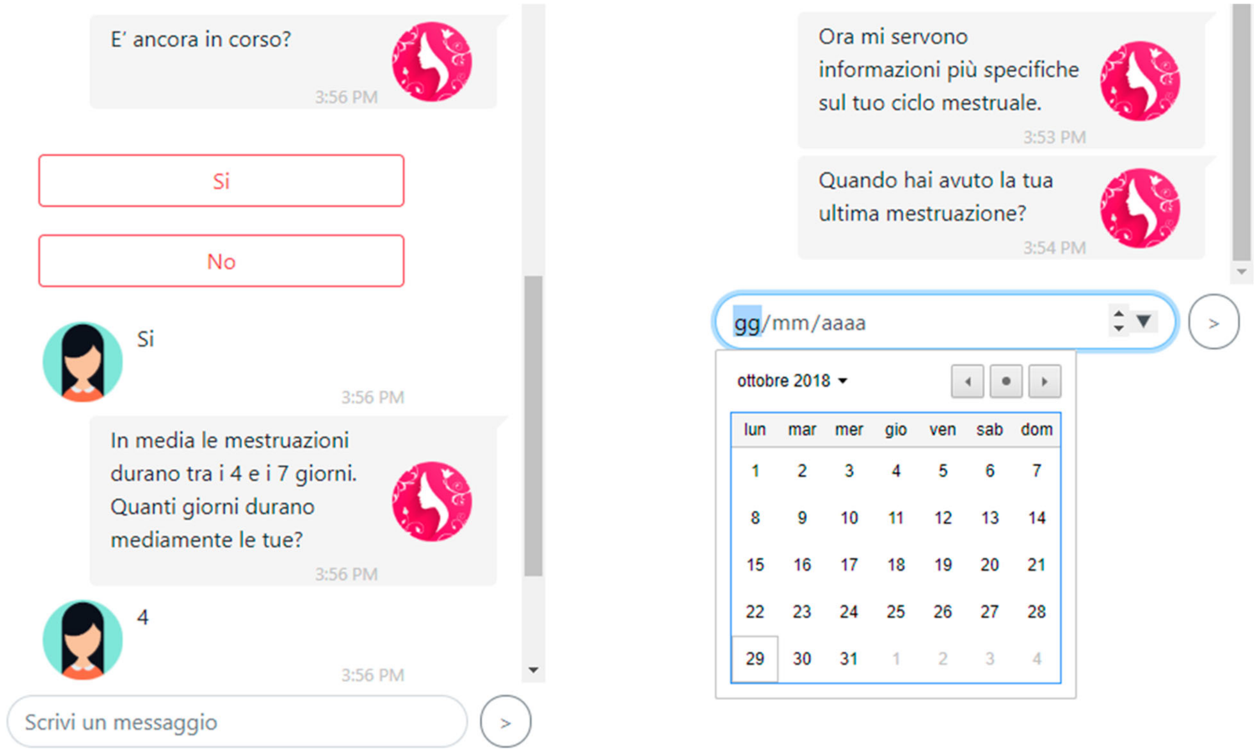


**Figure 3.** The 'LovApp' mobile app. The screen on the left shows the app asking preliminary information to the user; e.g. the duration of the period. The screen on the right shows the calendar feature that allows the user to tell the app the date of the last period.

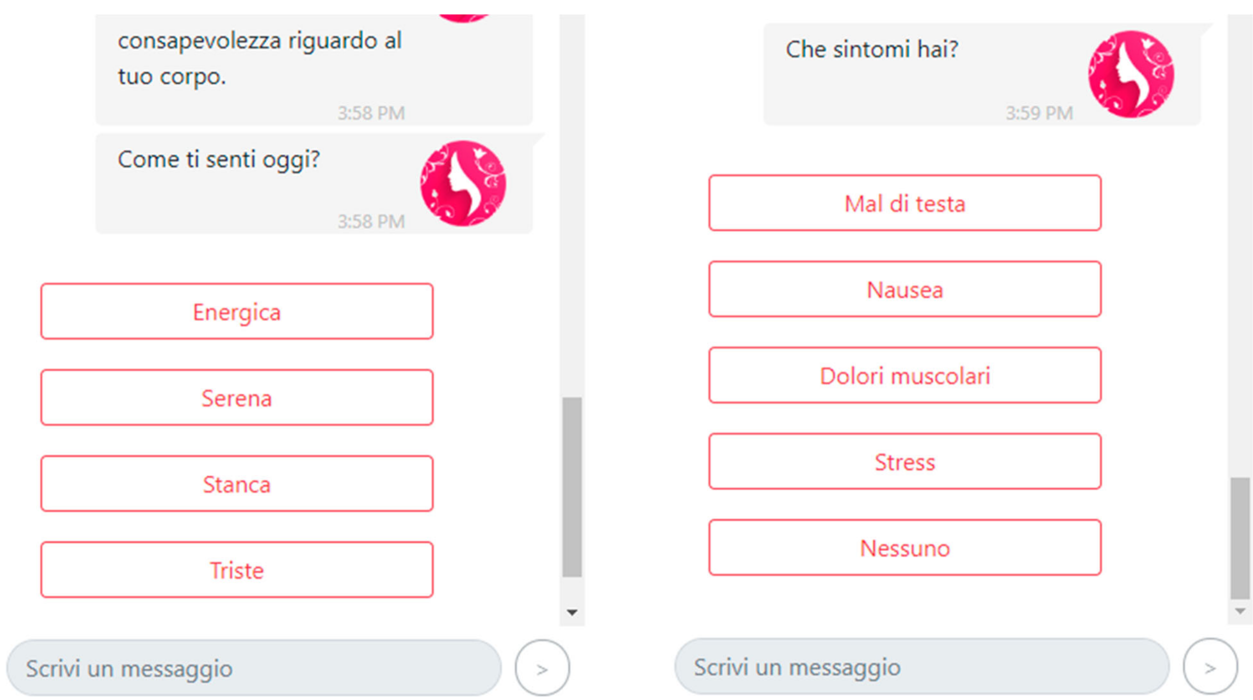


**Figure 4.** Further information entered in the app 'LovApp': on the left, the chatbot asks how the user feels, on the right the user can point out eventual symptoms.

user tests, the participants were asked to fill in an initial demographic questionnaire through which we collected information about their profiles.

#### 4.2.1. Healthcare

For testing the smart home domain apps, we involved 20 users (all female, due to the goals of the apps). Five of

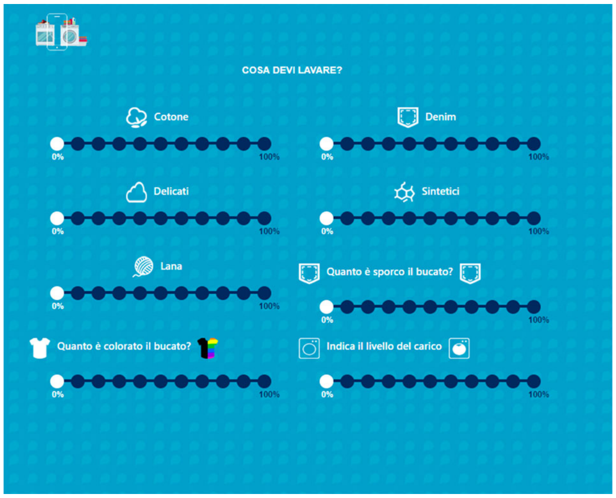


**Figure 5.** The 'SmartLiving' app asks information about the quality and quantity of clothes that the user needs to wash.


**Figure 6.** The user receives a suggestion that he/he can accept or decline.

them were aged between 18 and 24, eight between 25 and 34, and seven between 35 and 44. 45% of the users had a bachelor degree, 30% a master degree, 15% a high school diploma, 5% a Ph.D., and 5% a post-master degree. Eleven participants reported previous experience with a healthcare app (four 'almost every day', 4 'often but not daily', and three 'sometimes'). Those with previous experience reported that their feedback was positive; seven of them found the apps easy to use while four declared that it was not always easy to interact with them. The main types of healthcare apps used by the participants were designed for tracking of food, calories, physical activity, and drugs assumption.

#### *4.2.2. Smart home*

For the user tests on the smart home domain apps, we involved 20 users (10 females, 10 male). Five of them were aged between 18 and 24, ten between 25 and 34,

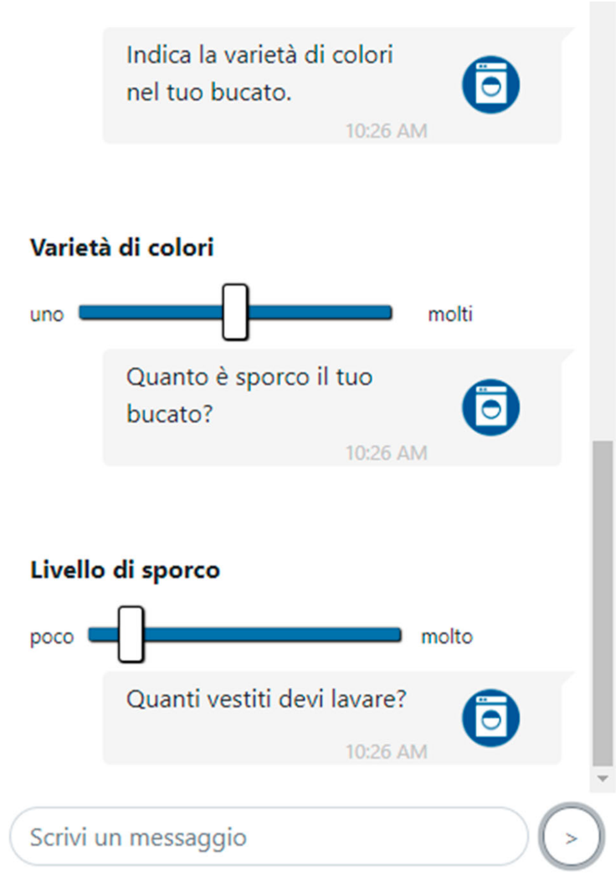


**Figure 7.** The user responds to a series of questions related to the quality and quantity of clothes loaded in the machine.

three between 35 and 44, one between 45 and 54, and one between 55 and 64. 40% of the users had a bachelor degree, 35% a high school diploma, and 25% a master degree. Five participants reported previous experience with a smart home app (one 'more than once a day', one 'almost every day', and three 'often but not daily'). Those with previous experience reported that their feedback was always positive and very easy and that they have used apps for remote control and monitoring of lights, heating, energy consumption, security cameras, and TV.

### *4.3. Settings*

The participants were invited to join the user test in a quiet room and asked to use the apps on a 10-inch Android-based tablet.

The user test protocol was designed by pairing cognitive and semiotic methods of usability and UX evaluation, with the aim of collecting significant information for identifying not only eventual usability and UX problems but also for highlighting the communication breakdowns that might take place during the interaction with the applications.

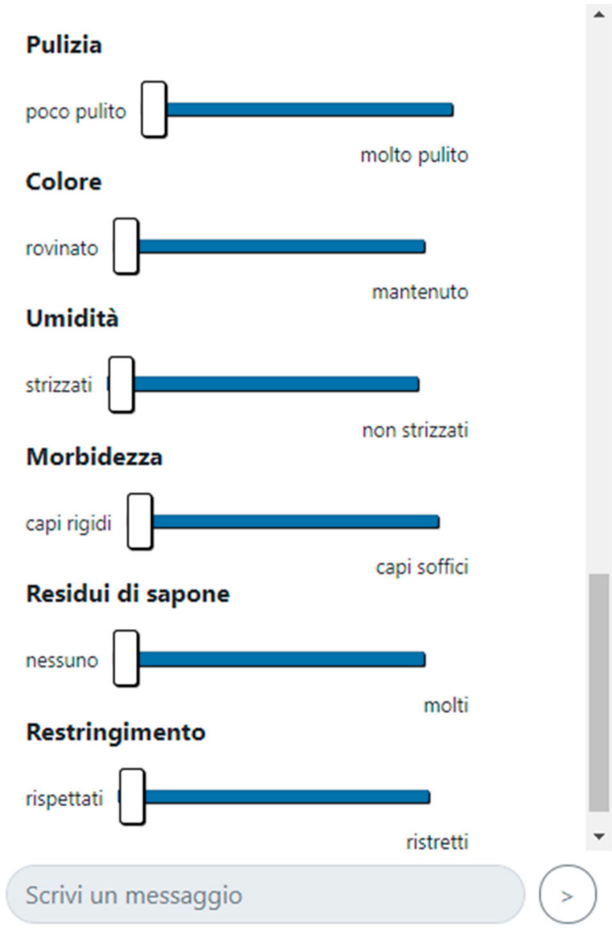


**Figure 8.** The 'SmartLiving Chatbot' app asks for feedback on the last suggestion provided.

### 4.4. Methodology

After the submission to the participants of the initial demographic questionnaire, the participants were asked to take part in a user test on the traditional application first and with the chatbot one then. The user test was scenario-based. During the test, the participants were asked to follow the think-aloud protocol, i.e. to interact with the apps and at the same time to verbalise their thoughts.

The participants always used the traditional application before using the chatbot by following the same scenario. Maintaining the same order did not affect the user experience since the chat-based interfaces and interaction strategies are completely different from the traditional GUI of mobile apps. In designing the conversation at the basis of the chatbot we did not simply migrate services from the mobile application to a conversational application. We changed the interaction protocol from a strategy based on free navigation to a fluid and realistic dialog between the user and the system.

At the end of each test, the participants filled in a usability questionnaire that was composed by three different sets of questions: a SUS (System Usability Scale) questionnaire (Brooke 1996), a CSUQ (Computer System Usability Questionnaire) (Lewis 1995) and a UEQ (User Experience Questionnaire) (Laugwitz, Held, and Schrepp 2008).

Finally, some open-ended questions were asked and the observations made by the participants were collected and analysed.

#### 4.4.1. User test: Scenario and CEM analysis

We used two scenarios, one for each domain: healthcare and smart home:

- Healthcare: You have always had a rather regular menstrual cycle but in the last few months, you have noticed delays, irregular cycles, and discomfort. Today you feel tired and decide to use a mobile app that allows you to record the days when you have the cycle, the basal temperature (which you often measure) and your symptoms. Use the app imagining that your menstrual cycle started on February 1st, 2019 and has an average duration of 4 days. Your flow is abundant, your basal temperature today is 37.20°, and you have nausea.
- Smart home: You went to live alone and among the various household appliances you bought, there is a new washing machine, a Rex Electrolux model RE 100. You are not an expert on laundry and they advise you to use a mobile app that offers advice on how to set up the washing machine. Use the app to get advice on how to wash cotton shirts and cotton pants, in various colours. Your basket is half-filled.

All user test sessions were directly observed by two researchers. This allowed applying the Communicability Evaluation Method (CEM) (Prates, de Souza, and Barbosa 2000) (De Souza and Leitão 2009), a Semiotic Engineering method for evaluating the communicability of an application. Specifically, the CEM method is focused on empirically studying the reception of the message sent by the designer to the user (by means of the system's interface). For CEM analysis, the two observers have to identify immediately during the test all hints of communicability breakdowns in the user's interaction.

After the test, the researchers have to associate tags to each communicability breakdown they identified. The process of tagging was done together by the researchers who discussed the outcomes of the test. For CEM tagging phase, the tags must be chosen among the 13 utterances proposed by De Souza and Leitão (2009): 'I give up.', 'Looks fine to me.', 'Thanks, but no, thanks.', 'I can do otherwise.', 'Where is it?', 'What happened?', 'What now?', 'Where am I?', 'Oops!', 'I can't do it this way.', 'What is this?', 'Help!', 'Why doesn't it?'.

During the next phase, called 'Interpretation', the researchers analyse the collected empirical data under four different perspectives: 1) Frequency and context of occurrence of each tag; 2) The existence of patterned sequences of tags; 3) The importance level of problems signalled by the occurrence of tags and patterns; and 4) The communicability issues that have caused the communicability breakdowns.

#### 4.4.2. SUS questionnaire

SUS (System Usability Scale) (Brooke 1996) is a very broadly used questionnaire, especially in industry, that can return reliable results even when administered to a small sample of users. It is composed of 10 statements with a five-point psychometric Likert scale (from 1 'Strongly disagree' to 5 'Strongly agree'). The SUS questionnaire uses alternatively positive (odd items) and negative (even items) wording.

When the results of the SUS questionnaire are back, the evaluators have to change the numbers as follows: for each of the odd-numbered questions, they have to subtract 1 from the score, while for each of the even numbered questions, they have to subtract the score from 5. The new values have then to be added and the result has to be multiplied by 2.5 (to shift the score to a 100-points scale – even if the result is not evaluated as a percentage).

According to SUS definition, the average score is 68.

#### 4.4.3. CSUQ questionnaire

CSUQ (Computer System Usability Questionnaire) (Lewis 1995) is a questionnaire developed by IBM and mostly focused on measuring the satisfaction in using the application or tool under evaluation. It consists of 19 statements with a five-point psychometric Likert scale (from 1 'Strongly disagree' to 5 'Strongly agree', with 3 as average score). Unlike SUS, the CSUQ items use only positive wording.

CSUQ has four internal subscales: System usefulness (SYSUSE), Information quality (INFOQUAL), Interface quality (INTERQUAL), and Overall (OVERALL). Their value ranges from 1 (worst) to 5 (best), therefore its average score is 2.5: SYSUSE is calculated as the average of the scores of questions 1–8 and question 19; INFOQUAL considers questions 9–15; INTERQUAL refers to questions 16–18, and OVERALL is the average of all questions' score.

#### 4.4.4. UEQ questionnaire

UEQ (User Experience Questionnaire) (Laugwitz, Held, and Schrepp 2008) permits to assess feelings, impressions, and attitudes that arise when the users use the application under evaluation. The UEQ questionnaire is made of 26 pairs of antithetic adjectives. The order of the adjectives is randomised: half of the items in the questionnaire has a positive adjective on the left side and the other half of the right side. The score ranges from −3 (worst) to + 3 (best) with 0 as the average score. The adjectives are related to six scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. Moreover, UEQ responses can be analysed for obtaining a more general result in terms of: Attractiveness, Pragmatic Quality, and Hedonic Quality. Pragmatic and Hedonic refer to the UX model by Hassenzahl, Schöbel, and Trautmann (2008) that assumes that people perceive products either as the ability to support the achievement of a specific goal (pragmatics) or the ability to support the individual/self-achievement, by making the users feeling competent.

#### 4.4.5. Open-ended questions

At the end of the questionnaires, the participants were asked to respond to some open-ended questions that regarded the experience of use they just made and the comparison with other similar products:

- Compared to other apps you know and use, how do you evaluate this experience of use?
- Could you please elaborate on this evaluation?
- Is there anything specific you would change in this app?
- Do you want to suggest any specific feature to be added to the app?
- Please, describe the app with one short sentence.

## 5. Results

In this Section, we report the results for each of the two domains. The discussion of the results is presented later, in Section 5.

### 5.1. Healthcare: 'Il Mio Calendario'

#### 5.1.1. User test – CEM

During the test, sixteen out of twenty participants had problems in finding the right features for recording the data suggested by the provided scenario. In fact, the 'Where is it' semiotic tag is the one that recurred more often (16 times). Other two tags appear just once each: 'What now?' and 'Oops!'. Given the scarcity of communicability breakdowns, it was not possible to detect patterns.

#### 5.1.2. SUS

The result of SUS analysis is slightly above the sufficiency: 68.87/100 (SD 17.38). All items received a fair score, with almost no difference between the results of positively and negatively worded items.

### 5.1.3. CSUQ

The results of the CSUQ questionnaire (see Figure 9) show that all internal subscales are above the average (i.e. 3 out of 5, where 5 is the best score). Specifically, the best result is reached with INTERQUAL, the subscale that measures the quality of the interface (4.05/5), while the lowest one is on INFOQUAL, quality of information (3.50/5) that is just slightly above the average.

The item that received the lowest score (2.75/5) is the one corresponding to the statement 'The system gave error messages that clearly told me how to fix problems'. This proves that the app is not easy to use because the error messages do not actually help the user in understanding how to overcome difficulties during the interaction. On the other hand, the statement that was best valued (4.2/5) is 'The interface of this system was pleasant', this means that the look-and-feel of the interface were appreciated despite the problems that users encountered during the test.

### 5.1.4. UEQ

The results of the UEQ questionnaires (see Figure 10) are all positive for 'Il Mio Calendario' app, however, the Stimulation scale reports the lowest value among the others (0.850/3). On the other hand, Efficiency scale is the one that scored the highest result (1.638/3). The results show that among the three qualities measured with UEQ, the one that appears to be weak is the Hedonic one (0.96/3), while Attractiveness and Pragmatic are higher but not particularly high.

### 5.1.5. Open-ended questions

Eleven participants out of 20 declared that they had previous experience with health apps. From their answers to the open-ended questions, almost everyone (10 out of 11) evaluated the user experience with 'Il Mio Calendario' app better than the ones they had with other similar apps. Specifically, the app was described as simpler, with an intuitive and nice interface, useful, and safe. When asked what they would change in the app, 5 out of 20 suggested to simplify the selection of the symptoms, 1 user responded that the interface should be made less childish, and another one suggested to improve the information architecture, while the other 13 users did not suggest any changes. To the question 'Do you want to suggest any specific feature to be added to the app?', just two proposals were made: introducing a search feature for finding the symptoms faster and easier, and adding push notifications when the menstrual cycle is expected to begin. The positive adjectives that were mostly used to describe the app were *useful*

| Items | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 3.75 | 3.90 | 3.40 | 3.55 | 3.70 | 3.70 | 3.75 | 3.55 | 2.75 | 3.15 | 3.55 | 3.70 | 4.10 | 3.55 | 3.70 | 4.20 | 3.85 | 4.10 | 3.85 |

| | |
|---|---|
| SYSUSE | 3.68 (SD 0.67) |
| INFOQUAL | 3.50 (SD 0.60) |
| INTERQUAL | 4.05 (SD 0.88) |
| OVERALL | 3.67 (SD 0.62) |

**Figure 9.** CSUQ scores for 'Il Mio Calendario' app.

| UEQ Scales | |
|---|---|
| Attractiveness | 1.383 (SD 1.02) |
| Perspicuity | 1.313 (SD 1.42) |
| Efficiency | 1.638 (SD 1.14) |
| Dependability | 1.425 (SD 0.77) |
| Stimulation | 0.850 (SD 1.19) |
| Novelty | 1.075 (SD 0.93) |

| UEQ Qualities | |
|---|---|
| Attractiveness | 1.38 |
| Pragmatic Quality | 1.46 |
| Hedonic Quality | 0.96 |

**Figure 10.** UEQ results for 'Il Mio Calendario' app.

(5), *nice* (3), *helpful* (3), *complete* (2), and *easy* (1), while the negative ones were *hard to learn* (2) and *messy* (1).

## 5.2. Healthcare: 'LovApp'

### 5.2.1. User test – CEM

During the user test, none of the participants encountered problems in using the app. In fact, the CEM analysis did not detect communicative breakdowns. As it will be better explained in Section 4.2.5, by answering the open-ended questions the users expressed their preference toward the chatbot-based app.

### 5.2.2. SUS

The result of SUS analysis reaches the value 79.87/100 (SD 9.20) that is well above the sufficiency. Also, in this case, the SUS results do not highlight specifically low or high scores among the various items.

### 5.2.3. CSUQ

Figure 11 shows that the internal subscales of the CSUQ questionnaire results all are above the average. The subscale with the highest score is SYSUSE (4.05/5), showing that the usefulness of the app is well recognised by the users. The INFOQUAL subscale that reaches the lowest result (3.50/5) still remains above the sufficiency.

The statement 'The system gave error messages that clearly told me how to fix problems.' is again the one that received the lowest score (2.95/5), while in this case, the one that reached the best result (4.65/5) is 'The organization of information on the system screens was clear.'.

### 5.2.4. UEQ

The UEQ results are illustrated in Figure 12. In the case of 'LovApp' app, Perspicuity reaches a high score (2.175/3) and also the other scales present positive results, especially Efficiency (1.688/3) and Attractiveness (1.608/3). Again, Stimulation appears to be the lowest scale (1.325/3). Moreover, the quality that reaches the highest score is the Pragmatic one (1.78/3) and the lowest score is obtained by the Hedonic one (1.36).

### 5.2.5. Open-ended questions

Seven out of the eleven users who had previous experience with health apps defined 'LovApp' as better than the other apps they knew and used, while 2 users described it as worse and 2 were not sure how to answer. Who found it better enjoyed the conversational interface because it makes the experience of using simpler and faster. The users who found the app worse than the others said that 'LovApp' is less nice and they do not like to have a conversation with the application. Among the suggestions received by the users, the most significant points were to add more symptoms, to provide the possibility of skipping the initial questions if the user feels

| Items | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 4.25 | 4.55 | 3.50 | 3.85 | 4.10 | 4.20 | 4.45 | 4.00 | 2.95 | 3.15 | 4.15 | 4.15 | 4.45 | 4.35 | 4.65 | 3.80 | 4.40 | 3.80 | 4.15 |

| | |
|---|---|
| SYSUSE | 4.12 (SD 0.44) |
| INFOQUAL | 3.98 (SD 0.48) |
| INTERQUAL | 4.00 (SD 0.57) |
| OVERALL | 4.05 (SD 0.41) |

**Figure 11.** CSUQ scores for 'LovApp' app.

| UEQ Scales | |
|---|---|
| Attractiveness | 1.608 (SD 0.76) |
| Perspicuity | 2.175 (SD 0.62) |
| Efficiency | 1.688 (SD 0.70) |
| Dependability | 1.488 (SD 0.80) |
| Stimulation | 1.325 (SD 0.89) |
| Novelty | 1.400 (SD 1.04) |

| UEQ Qualities | |
|---|---|
| Attractiveness | 1.61 |
| Pragmatic Quality | 1.78 |
| Hedonic Quality | 1.36 |

**Figure 12.** UEQ results for 'LovApp' app.

expert enough, to add a feature that reminds about the contraceptive pill, and to give the possibility of record intercourses. The positive adjectives used to describe 'LovApp' were *helpful* (5), *simple* (4), *useful* (4), and *guiding* (1), while the negative adjectives were *immature* (1) and *too simple* (1).

## 5.3. Smart home: 'SmartLiving'

### 5.3.1. User test – CEM

The results of the CEM analysis on the user tests point out that often participants experienced communicability breakdowns during the interaction with the app. In fact, three tags were identified several times by the observers: 'Help!' (8 times), 'What now?' (7 times), 'What's this?' (10 times).

Also, some patterns recurred: 'What now?' – > 'What's this?' – > 'Help!' (2 times), 'What now?' – > 'Help!' (2 times), 'What now?' – > 'What's this?' (2 times), and 'Help!' – > 'What's this?' (2 times).

### 5.3.2. SUS

The result of SUS analysis reaches the value 63.75/100 that is below the sufficiency (68/100 SD 14.88). All items reached modest results without particular low or high score to discuss.

### 5.3.3. CSUQ

The results of the CSUQ questionnaire (see Figure 13) show that all internal subscales are only just above the average (i.e. 3 out of 5, where 5 is the best score). The best result is reached with INTERQUAL, the subscale that measures the quality of the interface (3.98/5), while the lowest one is on INFOQUAL, quality of information (3.30/5).

The item corresponding to the statement 'The system gave error messages that clearly told me how to fix problems.' is still the one that reaches the lowest result (2.6/5), while the ones with the same best score (4.05/5) are 'The interface of this system was pleasant.' and 'This system has all the functions and capabilities I expect it to have.'

### 5.3.4. UEQ

The UEQ results related to the use of the 'SmartLiving' app are quite low (see Figure 14): the highest result is of 'Efficiency' scale (1.225/3) and the lowest is of 'Perspicuity' scale (0.975/3). Also, the general results are quite low, in fact, they range from 1.13/3 for 'Pragmatic Quality' and 1.06/3 for 'Hedonic Quality'.

### 5.3.5. Open-ended questions

Among the users who participated in the user test for the smart home domain, only 5 out of 20 declared that they had previous experience with smart home apps. Three out of five evaluated the user experience with 'Smart Living' app better than the ones they had with other similar apps. The app was described by them as useful, especially for saving time, money and energy. On the other hand, the users who described the app as worse than the others said that it was not clear how to describe the quality of the clothes loaded in the machine. One user suggested adding a feature that suggests the correct type of laundry detergent to use. The positive adjectives that were mostly used to describe the app were *helpful* (9), *useful* (5), and *innovative* (1), while the negative one was *immature* (1).

## 5.4. Smart home: 'SmartLiving Chatbot'

### 5.4.1. User test – CEM

The CEM analysis on the user test results pointed out that just one type of communicability breakdown was experienced during the interaction, i.e. 'What's this?' (8 times), therefore no pattern was detected in the user of 'SmartLiving Chatbot'.

### 5.4.2. SUS

The result of SUS analysis reaches the value 75.75/100 (SD 8.84) that is above the sufficiency.

| Items | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 3.70 | 3.70 | 2.95 | 3.40 | 3.80 | 3.70 | 3.60 | 3.80 | 2.60 | 3.15 | 2.75 | 3.70 | 3.80 | 3.85 | 3.25 | 4.05 | 3.85 | 4.05 | 3.70 |

| | |
|---|---|
| SYSUSE | 3.59 (SD 0.72) |
| INFOQUAL | 3.30 (SD 0.49) |
| INTERQUAL | 3.98 (SD 0.57) |
| OVERALL | 3.55 (SD 0.54) |

**Figure 13.** CSUQ scores for 'SmartLiving' app.

| UEQ Scales | |
|---|---|
| Attractiveness | 1.100 (SD 0.91) |
| Perspicuity | 0.975 (SD 1.16) |
| Efficiency | 1.225 (SD 1.05) |
| Dependability | 1.188 (SD 0.96) |
| Stimulation | 1.025 (SD 0.94) |
| Novelty | 1.100 (SD 0.78) |

| Qualities | |
|---|---|
| Attractiveness | 1.10 |
| Pragmatic Quality | 1.13 |
| Hedonic Quality | 1.06 |

**Figure 14.** UEQ results for 'SmartLiving' app.

### 5.4.3. CSUQ

The results of the CSUQ questionnaire presented in Figure 15, show that all internal subscales are only just above the value 3 that is the average score. The best result is reached with INTERQUAL, the subscale that measures the quality of the interface (3.95/5), while the one with the lowest score is INFOQUAL, quality of information (3.73/5).

Like the other cases, the lowest score (2.75/5) is the one associated with the statement 'The system gave error messages that clearly told me how to fix problems.'. The statement 'The information provided for the system was easy to understand.' received the best average score (4.35/5).

### 5.4.4. UEQ

Figure 16 shows the results of the UEQ questionnaire. It shows that the only scale that reaches a score above 2 is 'Perspicuity' (2.100/3), while the lowest scores are those reached by 'Efficiency' and 'Stimulation' (both 1.338/3). Among the qualities, the one that shows the most positive result is 'Pragmatic Quality' (1.63/3), while 'Attractiveness' is the worst (1.43/3).

### 5.4.5. Open-ended questions

Four out five users who had previous experiences with smart home apps declared that 'SmartLiving Chatbot' app is better than the ones they knew and used (one preferred not to answer to the question). Specifically, the conversational interface was appreciated and described as easy to use, clear, and fast. As a future to be added, the users suggested introducing the possibility of returning to the beginning of the conversation (a sort of 'Home' button) and to introduce the suggestion of the correct detergent and additives to use. The main positive

| Items | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 4.05 | 4.25 | 3.40 | 3.55 | 3.95 | 4.00 | 4.20 | 3.90 | 2.75 | 2.85 | 3.75 | 3.95 | 4.35 | 4.30 | 4.15 | 3.85 | 4.00 | 4.00 | 4.05 |

| | |
|---|---|
| SYSUSE | 3.93 (SD 0.44) |
| INFOQUAL | 3.73 (SD 0.36) |
| INTERQUAL | 3.95 (SD 0.46) |
| OVERALL | 3.86 (SD 0.34) |

**Figure 15.** CSUQ scores for 'SmartLiving Chatbot' app.

| UEQ Scales | |
|---|---|
| Attractiveness | 1.425 (SD 0.76) |
| Perspicuity | 2.100 (SD 0,52) |
| Efficiency | 1.338 (SD 1.01) |
| Dependability | 1.463 (SD 0.50) |
| Stimulation | 1.338 (SD 0.98) |
| Novelty | 1.775 (SD 1.04) |

| Qualities | |
|---|---|
| Attractiveness | 1.43 |
| Pragmatic Quality | 1.63 |
| Hedonic Quality | 1.56 |

**Figure 16.** UEQ results for 'SmartLiving Chatbot' app.

adjectives used to describe 'SmartLiving Chatbot' are *helpful* (10), *easy* (3), *innovative* (2), *simple* (2), *useful* (1), and *guiding* (1).

## 6. Discussion

From SUS questionnaire results (see Figure 17), the chatbot-based apps appear to be slightly preferred by the users in both application domains (healthcare and smart home) but the mean values do not present statistically significant differences.

The results obtained by the chatbot-based apps through the CSUQ questionnaire is better for SYSUSE, INFOQUAL, and OVERALL. For what concerns the INTERQUAL (i.e. the quality of the interface) the results are very close both for the apps of the healthcare domain (Figure 18) and for those of the smart home domain (Figure 19). But also, in this case, the mean values do not present statistically significant differences. This may suggest further studies to investigate differences between usability and user experience testing results. In fact, where the usability gives us not an appreciable results user experience and communicability analysis can provide more useful feedbacks in using chatbots.

The interesting results are those that come from the UEQ questionnaire, i.e. the evaluation of the user

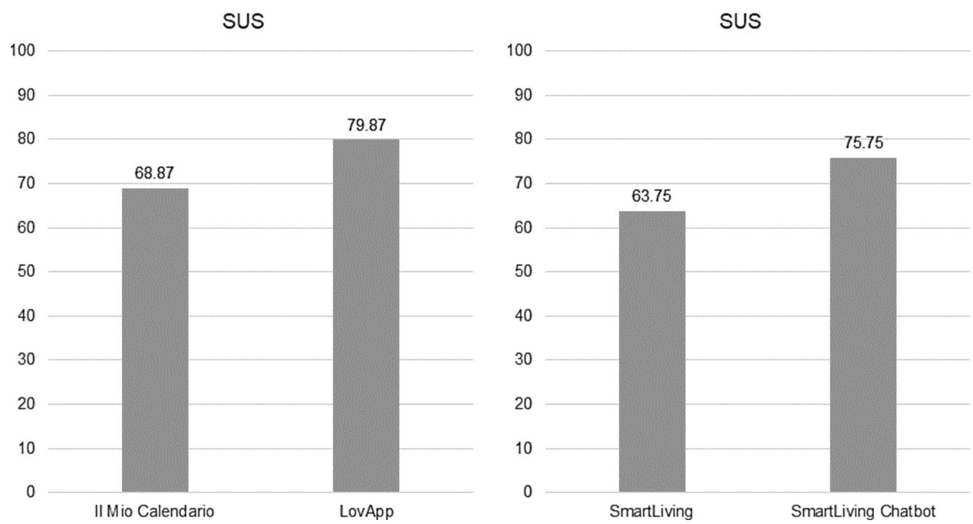


**Figure 17.** The comparison of the SUS results for the healthcare (left) and the smart home (right) domains.

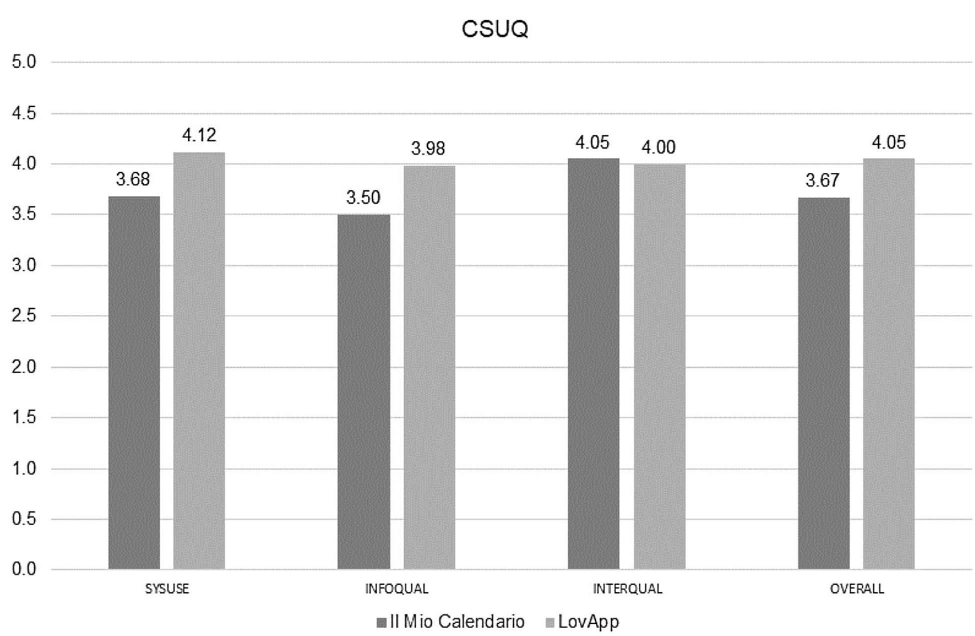


**Figure 18.** The comparison of the CSUQ results for the healthcare domain.

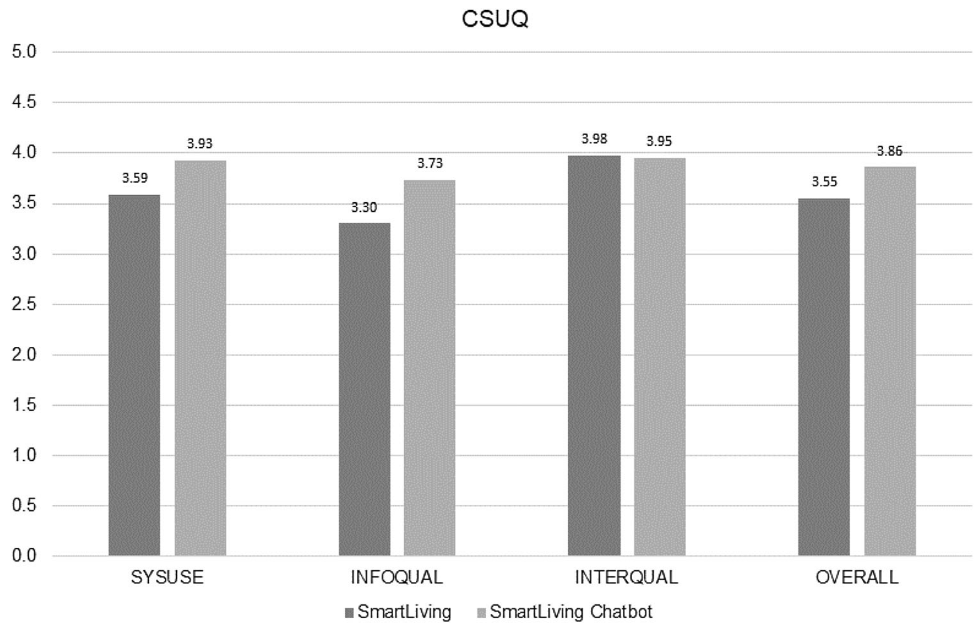


**Figure 19.** The comparison of the CSUQ results for the smart home domain.

experience. Figure 20 shows how the experience with the chatbot-based app is almost always better than the one with the traditional interface. Especially the results of the scales Attractiveness, Perspicuity, Simulation, and Novelty show the difference even if the mean values do not present statistically significant differences.

Figure 21 groups the results into UEQ Qualities and highlights how the conversational interfaces result better for Attractiveness, Pragmatic Quality, and also for Hedonic Quality.

In both studies, the responses to the open-ended questions and final users' comments were coded using thematic analysis (Marshall and Rossman 2014) in particular for what concerns the use of conversational interfaces with respect to the use of traditional interfaces. The categories and codes are listed in Figure 22 with sample quotes from the data. The responses highlighted the issues involving the ease-of-use, helpful and innovative approach of the chatbot interface and its future use. The participants indicated that the chatbot is easy to understand and apply. Many of the participants expressed an interest in investigating this new interaction strategy in future use. The participants also highlighted some limitations related to the fact the chatbot is a prototype of a system that needs a longer process of reflection, study, and analysis. In particular, they indicated that daily use of the chatbot could be boring and, especially for healthcare data monitoring, a daily

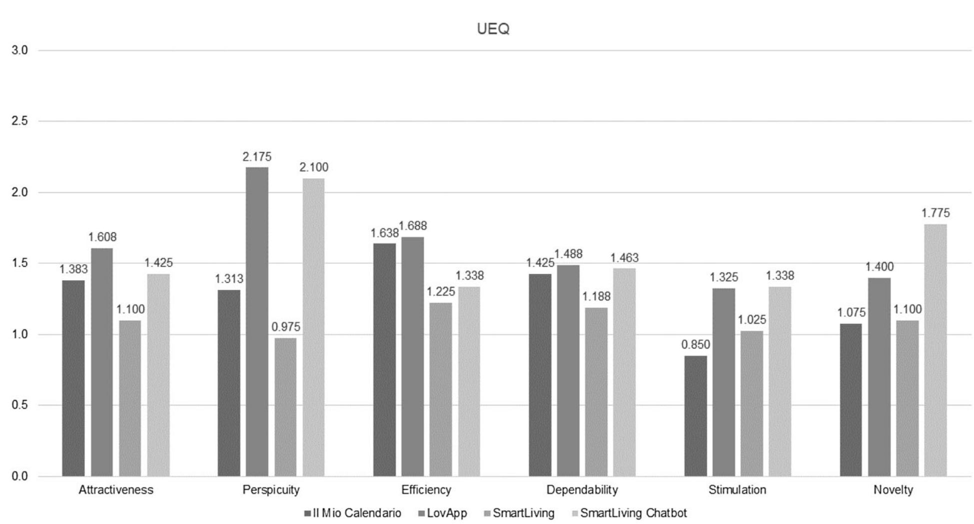


**Figure 20.** The comparison of the UEQ results for both the domains.

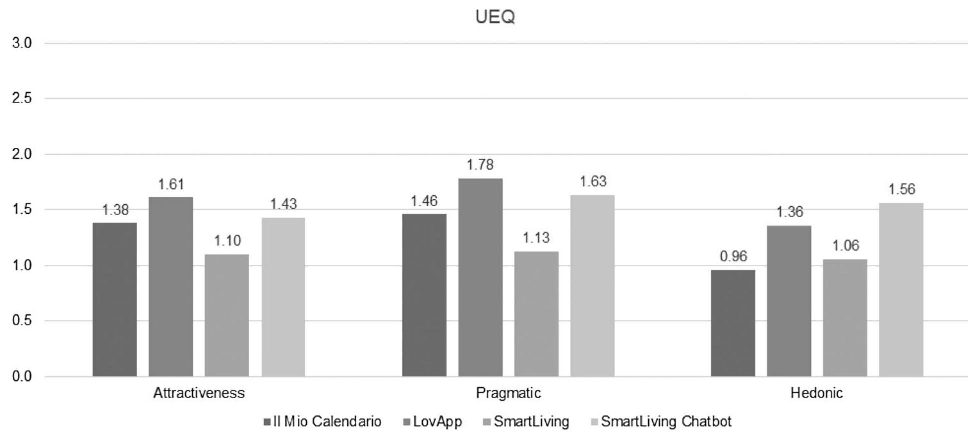


**Figure 21.** The comparison of the UEQ Qualities for both the domains.

| Code | Example quotes |
|---|---|
| (+) Easy to learn | • *"The chatbot is very simple for the first uses"*<br>• *"It is easy to use because it provides an appearance of human contact"*<br>• *"The bot is useful for those who has a poor attitude in using digital systems "* |
| (+) Helpful approach | • *"The dialogue with the bot improves the use experience"*<br>• *"The possibility of conversation makes data entry easier"*<br>• *"The bot seems to me more immediate, complete and searchable"* |
| (+) Future use | • *"It would be interesting to use it again, it seems to talk to an expert who can help you"*<br>• *"I like dialogue but it's binding I would like to try it for more time"*<br>• *I would like to try the chatbot better in the future because I'm interested in the idea of communicating via messages like in Whatapps"* |
| (-) Immaturity | • *"... at present it seems a little bizarre"*<br>• *"Some problems have been detected during the registration phase in the chatbot"*<br>• *"...in the chatbot I can find all features I need"* |
| (-) Boring system | • *"...however, over time the interaction could become tedious"*<br>• *"The bot is useful even if it risks becoming boring to wait for all the questions if you already know where to click and how to proceed"*<br>• *"I fear the bot may be intrusive"* |

**Figure 22.** Summary of the codes with sample quotes.

conversation could be tiresome and lead to annoying notifications. To check these issues, further studies are necessary specifically for monitoring a daily use of conversational agents in different contexts of use. Nevertheless, as a preliminary result, we can claim that integrating the chatbot with traditional GUI elements (such as buttons, lists, menus, sliders) appears to be a promising solution for improving the communicability level of the systems. These results are very much in line with the qualitative feedback presented in previous Sections.

The lesson we can learn from this study concerns the benefits chatbots offer through the implementation of a conversational interface that can help users to improve response time and their satisfaction and ensure user retention. The switch from a traditional GUI to a dialog-based system can favour users with a poor attitude in using technology. Instead of getting lost in interfaces that users have to explore and understand for solving their problems, a chatbot acts like a partner who helps you to find out specific kinds of information or to discuss specific issues. Specifically, in managing sensible information (such as in the health domain) or in dealing with and control IoT data and applications (such as in the smart home domain), conversational interfaces can provide users with a pro-active but not intrusive approach. As a partner, it is important to be proactive without be intrusive, and be friendly and inviting while still providing information that users need to hear. Sentiment analysis and machine learning techniques are clear examples of current strategies that it is possible to use for getting more and more intelligent chatbots.

In these contexts of use, conversational interfaces can help to deal with issues related to social inclusion and digital divide that lead to a failure in handling technology or to a poor attitude in using digital systems for solving also simple problems.

### *6.1. Limitations of the study*

We can identify three limitations of our study:

1. The small number of domains tested: Further development of this study should consider more than just two application domains. This study considered the healthcare and the smart home domains because they are simple, straightforward use cases where the goal is well defined, and a user's intent is easily identifiable and so they are really applicable for testing conversational user interfaces. Moreover, they stemmed from the projects that were specifically focused on these two domains.
2. The number of participants: Even though the size of the user group allowed us to reach interesting results, further development of the study would be considering the involvement of more users.
3. The duration of the user experience: The tests we performed were carried out in a limited amount of time (less than one hour for each participant). It would be interesting to observe/monitor the use of the apps over a much longer period of time to find out if the user experience improves or not during the time. Moreover, in such a long-duration study, it would be possible to study how the users are (or are not) engaged by the apps and to evaluate the difference of such engagement with both traditional and conversational interfaces.

## 7. Conclusions

Conversational interfaces are designed to respond in a way similar to what happens when another human being is involved in a conversation. It differs from the activity of instructing insofar as it encompasses a two-way communication process, with the system acting like a partner rather than a machine that obeys orders. It has been most commonly used for applications where the user needs to find out specific kinds of information or wants to discuss issues.

In this paper, we presented a study able to highlight the main benefits of developing a conceptual model that uses a conversational style of interaction is that it allows people to communicate with a system in a way that is familiar to them.

After conducting two comparison tests, differences have been found between the chatbots and traditional apps. In both the application domains, for usability and UX the chatbot application appears to be better than the traditional one. It is necessary to underline that the chatbot prototype we implemented is slower than the traditional app because it has a delay between a message and another. This delay is due to two reasons. First of all, while in a conversational interface at every step of the conversation there is a network communication between the client and the conversational agent and the UI elements are dynamically loaded into the client interface, in a traditional app, the UI is loaded immediately and only once. Secondly, by using a chatbot we can keep the user's pace. In detail, we can delay messages to avoid the risk of overloading users with fast and frequent messages (for example in the case of elderly people). This is also true when users are not familiar with the use of a chatbot. On the contrary, we can avoid taking the dialog longer than smarter users need by providing a quick cross talk (for example in the case of teenagers or younger adults). This because a slower performance in using the chatbot app may have negatively influenced the level of appreciation of the bot itself.

While in the healthcare case there were no communicability breakdowns, in the smart home one the CEM analysis was better for the chatbot application. Even if the chatbot results are below the maximum possible score, they are indeed better than the ones of the traditional applications. The problems detected with the chatbot could be explained as follows: For the smart home, a problem that may have induced negative results in the appreciation of the smart home application is the function of the application itself. Users did not really like using an app to receive wash tips. Paraphrasing what they said, the reaction was: 'Surely it is something new but I would never use it'. So the novelty in this sense is not in contrast with a low level of satisfaction; for healthcare, the possibility of feeling more comfortable with a bot does not imply the need to an app that controls health status. An aspect to take into consideration that may be a motivation for this result is that users involved do not have problems related to their menstrual cycle (i.e. problems of fertility, pregnancy research, etc.) so they probably do not feel the need to face these topics in a personal way.

In conclusion, we can say that for simple tasks with a clear goal (problem-based), a chatbot experience can be a promising solution for improving the communication between users and systems. Even if there are still technical challenges to overcome, a chatbot can offer a more conversational and so more familiar communication strategy. In order to demonstrate it, in the future we will validate these conversational interfaces with experiments in real contexts of use. Specifically, situations where users have to deal with daily problems for interfacing with smart home devices or situations where users with health problems need support for monitoring their biometric data or diet behaviour. This will provide a more significant assessment of the approach.

## Notes

1. https://play.google.com/store/apps/details?id=com.popularapp.periodcalendar&hl=it
2. www.draw.io/

## Acknowledgments

This work is partially supported by two projects funded by Regione Lombardia: I-VITAE New Pathways for Life and Social appliances for Industry 4.0 - EaSy 4.0 Smart Living. The authors want to thank Carlo Cassanelli for the collaboration in developing the chatbots used for this study.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

*Stefano Valtolina* http://orcid.org/0000-0003-1949-2992
*Barbara Rita Barricelli* http://orcid.org/0000-0001-9575-5542

## References

Andersen, P. B. 1997. *A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems*. New York, NY, USA: Cambridge University Press.

Ardito, C., B. R. Barricelli, P. Buono, M. F. Costabile, A. Piccinno, S. Valtolina, and L. Zhu. 2011. LNCS Volume 6765, Issue PART 1, 2011, Pages 3–11.

Barricelli, B. R., G. Fischer, D. Fogli, A. Mørch, A. Piccinno, and S. Valtolina. 2016. "Cultures of Participation in the Digital age: From "Have to" to "Want to" Participate." In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*, 1–3. New York, NY, USA: Association for Computing Machinery doi:10.1145/2971485.2987668

Barricelli, B. R., and S. Valtolina. 2017. "A Visual Language and Interactive System for end-User Development of Internet of Things Ecosystems." *Journal of Visual Languages and Computing* 40: 1–19.

Boratto, L., S. Carta, F. Mulas, and P. Pilloni. 2017. "An e-Coaching Ecosystem: Design and Effectiveness Analysis of the Engagement of Remote Coaching on Athletes." *Personal and Ubiquitous Computing* 21 (4): 689–704.

Brandtzaeg, P. B., and A. Følstad. 2017. "Why People use Chatbots." In *International Conference on Internet Science*, 377–392. Cham: Springer.

Braun, D., A. Hernandez-Mendez, F. Matthes, and M. Langen. 2017. "Evaluating Natural Language Understanding Services for Conversational Question Answering Systems." In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 174–185. Association for Computational Linguistics.

Brooke, J. 1996. "SUS: A Quick and Dirty Usability Scale." In *Usability Evaluation in Industry*, edited by P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, 189–194. London, UK: London: Taylor and Francis.

Cathleen, W., R. John, L. Clayton, and P. Peter. 1994. "The Cognitive Walkthrough Method: a Practitioner's Guide." In *Usability Inspection Methods*, 105–140. New York, USA: John Wiley & Sons Inc.

Chai, J. Y., M. Budzikowska, V. Horvath, N. Nicolov, N. Kambhatla, and W. Zadrozny. 2001. Natural language sales assistant-a web-based dialog system for on-line sales. In Iaai (pp. 19–26).

«Chatbots Market Research Report- Global Forecast 2023,» [Online]. Accessed 15 March 2018. https://www.marketresearchfuture.com/reports/chatbots-market-2981.

Chourabi, H., T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, and H. J. Scholl. 2012. "Understanding Smart Cities: An Integrative Framework." In *2012 45th Hawaii int. Conference on System Sciences*, 2289–2297. Maui, HI, USA: IEEE.

Colby, K. M. 1999. "Human-computer Conversation in a Cognitive Therapy Program." In *Machine Conversations*, 9–19. Boston, MA: Springer.

De Souza, C. S. 2005. *The Semiotic Engineering of Human-Computer Interaction*. Cambridge, MA, USA: MIT press.

De Souza, C. S., and C. F. Leitão. 2009. *Semiotic Engineering Methods for Scientific Research in HCI*. San Rafael: Morgan and Claypool Publishers.

De Souza, C. S., C. F. Leitão, R. O. Prates, and E. J. da Silva. 2006. "The Semiotic Inspection Method." In *Proc. of the 7th Brazilian Symposium of Human Factors on Computer Systems*, 148–157. Porto Alegre: SBC.

Dirican, C. 2015. " The Impacts of Robotics." *Artificial Intelligence On Business and Econom-ics. Social and Behavioral Sciences* 195: 564–573. doi:10.1016/j.sbspro.2015.06.134.

Dirks, S., C. Bühler, and C. Edler. 2018. "Digital Inclusion Through Accessible Technologies." In *International Conference on Computers Helping People with Special Needs*, 435–438. Cham: Springer.

Fogg, B. J. 2009. "A Behavior Model for Persuasive Design." In *Proceedings of the 4th International Conference on Persuasive Technology*, 40–40. Claremont, California, USA: ACM.

Følstad, A., and P. B. Brandtzæg. 2017. " Chatbots and the New World of HCI." *Interactions* 24 (4): 38–42.

Gomes da Silva, P., and A. Dix. 2007. "Usability - Not as we Know it!." In *Proceedings of HCI 2007*, 1–4. Plymouth , United Kingdom: British Computer Society.

Grasso, F., A. Cawsey, and R. Jones. 2000. "Dialectical Argumentation to Solve Conflicts in Advice Giving: a Case Study in the Promotion of Healthy Nutrition." *International Journal of Human-Computer Studies* 53 (6): 1077–1115.

Greenberg, S., and W. Buxton. 2008. "Usability Evaluation Considered Harmful (Some of the Time)." In *Proc. of CHI 2008*, 111–120. New York, NY, USA: ACM.

Hassenzahl, M., M. Schöbel, and T. Trautmann. 2008. "How Motivational Orientation Influences the Evaluation and Choice of Hedonic and Pragmatic Interactive Products: The Role of Regulatory Focus." *Interacting with Computers* 20 (4-5): 473–479.

James, G. 2016. «A Complete Guide to Chatbots» [Online]. Accessed 21 February 2018. http://www.garethjames.net/completeguidechatbots/.

Jenkins, M. C., R. Churchill, S. Cox, and D. Smith. 2007. "Analysis of User Interaction with Service Oriented Chatbot Systems." In *International Conference on Human-Computer Interaction*, 76–83. Berlin Heidelberg, Germany: Springer Berlin Heidelberg.

Jongeling, R., S. Datta, and A. Serebrenik. 2015. "Choosing Your Weapons: On Sentiment Analysis Tools for Software Engineering Research." In *2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, 531–535. Washington, DC, USA: IEEE.

Khiyal, M. S. H., A. Khan, and E. Shehzadi. 2009. "SMS Based Wireless Home Appliance Control System (HACS) for Automating Appliances and Security." *Issues in Informing Science & Information Technology* 6: 887–894.

Laugwitz, B., T. Held, and M. Schrepp. 2008. "Construction and Evaluation of a User Experience Questionnaire." In *USAB 2008, LNCS 5298*, 63–76. Berlin, Germany: Springer Berlin Heidelberg.

Letzter, R. 2016. Ibm's brilliant ai just helped teach a grad-level college course. Accessed 24 June 2017. http://uk.businessinsider.com/watson-ai-became-a-teaching-assistant-2016-5?international=truer=UK&IR=T.

Lewis, J. R. 1995. "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use." *International Journal of Human-Computer Interaction* 7 (1): 57–78.

Lindgren, H., E. Guerrero, and R. Janols. 2017. "Personalised Persuasive Coaching to Increase Older Adults' Physical and Social Activities: a Motivational Model." In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, 170–182. Cham: Springer.

Marcante, A., and P. Mussio. 2006. *Electronic Interactive Documents and Knowledge Enhancing: a Semiotic Approach*. UC Berkeley: The Document Academy.

Marshall, C., and R. Rossman. 2014. *Designing Qualitative Research*. 2nd Edition. New York, NY, USA: Sage Publications.

MarutiTechlabs. 2017. Accessed 7 November 2018. https://www.facebook.com/Customer-Support-Bot-1857341381220252/.

M. C. Agency. 2016. «Chatbots: a consumer research study».

Morinaga, S., K. Yamanishi, K. Tateishi, and T. Fukushima. 2002. "Mining Product Reputations on the web." In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 341–349. New York, NY, USA: ACM.

Nielsen, J., and R. L. Mack. 1994. *Usability Inspection Methods*. New York: John Wiley & Sons.

Nocera, J. A., B. R. Barricelli, A. Lopes, P. Campos, and T. Clemmensen. 2015. "Preface." In *IFIP Advances in Information and Communication Technology*, 468. New York LLC: Springer.

Norman, D. 1992. "Why Interfaces Don't Work." In *The Art of Human-Computer Design*, edited by Brenda Laurel, 209–219. Boston, MA, USA: Addison-Wesley Publishing Company, Inc.

Oinas-Kukkonen, H., and M. Harjumaa. 2018. "Persuasive Systems Design: key Issues, Process Model and System Features." In *Routledge Handbook of Policy Design*, 105–123. New York, NY, USA: Routledge.

Prates, R. O., C. S. de Souza, and S. D. J. Barbosa. 2000. "A Method for Evaluating the Communicability of User Interfaces." *ACM Interactions* 7 (1): 31–38.

Robino, G. 2018. «Il 2018 è l'anno delle interfacce vocali?» ConvComp.it, [Online]. Accessed 15 March 2018. https://convcomp.it/il-2018-%C3%A8-lanno-delle-voice-user-interfaces-3134ca939497.

Shawar, B. A., and E. Atwell. 2007. "Different Measurements Metrics to Evaluate a Chatbot System." In *Proceedings of the Workshop on Bridging the gap: Academic and Industrial Research in Dialog Technologies*, 89–96. Stroudsburg, PA, USA: Association for Computational Linguistics.

Shawar, A., E. Atwell, and A. Roberts. 2005. "Faqchat as in Information Retrieval System." In *Human Language Technologies as a Challenge for Computer Science and Linguistics: Proceedings of the 2nd Language and Technology Conference*, 274–278. Poznan, Poland: Wydawnictwo Poznanskie.

Tan, K. K., T. H. Lee, and C. Y. Soh. 2002. "Internet-based Monitoring of Distributed Control Systems-An Undergraduate Experiment." *IEEE Transactions on Education* 45 (2): 128–134.

Tatai, G., A. Csordàs, À Kiss, A. Szalò, and L. Laufer. 2003. "Happy Chatbot, Happy User." In *International Workshop on Intelligent Virtual Agents*, 5–12. Berlin Heidelberg, Germany: Springer Berlin Heidelberg.

Thompson, A., and E. Kemp. 2009. "Web 2.0: Extending the Framework for Heuristic Evaluation." In *Proc. of CHINZ '09*, 29–36. New York: ACM Press.

Tondl, L. 1981. Problems of Semantics, Reidel Publishing.

Torchiani, G. 2018. «Chatbot: cosa sono e cosa c'è dietro il loro successo,» Al4Business, [Online]. Accessed 15 March 2018. https://www.ai4business.it/robotica/chatbot/chatbot-cosa-cosa-ce-dietro-successo/.

Turney, P. D., and M. L. Littman. 2003. "Measuring Praise and Criticism: Inference of Semantic Orientation From Association." *ACM Transactions on Information Systems (TOIS* 21 (4): 315–346.

Valtolina, S., B. R. Barricelli, and Y. Dittrich. 2012. "Participatory Knowledge-Management Design: a Semiotic Approach." *Journal of Visual Languages and Computing* 23 (2): 103–115.

Venkatesh, A., C. Khatri, A. Ram, F. Guo, R. Gabriel, A. Nagar, and R. Goel. 2018. "On Evaluating and Comparing Conversational Agents." *arXiv Preprint ArXiv:1801.03625* 4: 60–68.

W.H.O. 1995. «Health promotion: Ottawa charter».

Wilson, A. 2018. «Chatbot security concerns» StaySafeOnline, 5 March 2018. [Online]. Accessed 08 October 2018. https://staysafeonline.org/blog/chatbot-security-concerns/.

Yan, M., and H. Shi. 2013. "Smart Living Using Bluetooth-Based Android Smartphone." *International Journal of Wireless & Mobile Networks* 5 (1): 65.

Zhu, L., B. R. Barricelli, and C. Iacob. 2011. "A Meta-Design Model for Creative Distributed Collaborative Design." *International Journal of Distributed Systems and Technologies* 2 (4): 1–16.

Zhu, L., P. Mussio, B. R. Barricelli, and C. Iacob. 2010. A habitable space for supporting creative collaboration. 2010 International Symposium on Collaborative Technologies and Systems, CTS 2010, art. no. 5478455, pp. 617–622.

Zue, V. W., and J. R. Glass. 2000. "Conversational Interfaces: Advances and Challenges." *Proceedings of the IEEE* 88 (8): 1166–1180.