**RESEARCH ARTICLE**

# Users' experiences with chatbots: findings from a questionnaire study

**Asbjørn Følstad**[1] · **Petter Bae Brandtzaeg**[1,2]

## Abstract

For chatbots to be broadly adopted by users, it is critical that they are experienced as useful and pleasurable. While there is an emerging body of research concerning user uptake and use of chatbots, there is a lack of theoretically grounded studies detailing what constitutes good or poor chatbot user experiences. In this paper, we present findings from a questionnaire study involving more than 200 chatbot users who reported on episodes of chatbot use that they found particularly satisfactory or frustrating. The user reports were analysed with basis in theory on user experience, with particular concern for pragmatic and hedonic attributes. We found that pragmatic attributes such as efficient assistance (positive) and problems with interpretation (negative) were important elements in user reports of satisfactory and frustrating episodes. Hedonic attributes such as entertainment value (positive) and strange and rude responses (negative) were also frequently mentioned. Older participants tended to report on pragmatic attributes more often, whereas younger participants tended to report on hedonic attributes more often. Drawing on the findings, we propose four high-level lessons learnt that may benefit chatbot service providers, and we suggest relevant future research.

**Keywords** Chatbots · User experience · Questionnaire study

## Introduction

Chatbots are machine agents with which users interact through natural language dialogue, by text or voice [4]. The conversational character of chatbots enables new and potentially more convenient and personal ways to access content and services [20]. Chatbots are employed across a range of application areas, such as customer service [38], health [16], education [19] and personal assistance [37]. While chatbots were initially developed to mimic human conversation [47], often in the form of social chatter, current chatbots are typically task-oriented, enabling their users to achieve specific goals or outcomes [41]. However, current task-oriented chatbots are typically also designed in consideration of the emotional or social aspects of conversational interaction [22]. Furthermore, chatbots developed purely for social interaction or relationship formation still exist as a distinct type of chatbot [42].

Chatbots are seen as promising by service providers [9]. They allow for efficient interactions with users on private messaging platforms and in virtual assistants, and interactions are familiar and low threshold [20]. For users, chatbots are considered a potentially efficient and enjoyable means of accessing content and services [5, 13]. Conversational interaction with machine agents is also becoming more commonplace due to the substantial uptake of voice assistants [32]. However, the full potential of chatbots is still not realized, in part due to challenges associated with changing user needs and motivations [4].

To enable a broad popular uptake of chatbots as a user interface to content and services, chatbots should provide their users with valuable and pleasing experiences to increase the likelihood that they will become regular chatbot users and increase their reliance on chatbots [2]. In short, chatbots need to be developed and designed with the aim of strengthening user experience.

Developing chatbots that strengthen user experience entails a number of research and innovation challenges pertaining to, for example, underlying technological enablers for natural language processing and context recognition as well as interaction design. At the same time, to improve chatbot user experience in the future, it is important to have

✉ Asbjørn Følstad
asf@sintef.no

1 SINTEF, Oslo, Norway

2 University of Oslo, Oslo, Norway

insight into what constitutes good and poor chatbot user experience today. Such insight may be valuable to guide future chatbot research and development.

There is an emerging body of research exploring how chatbots are perceived and used. However, there is a lack of knowledge regarding what users see as particularly good or particularly poor chatbot user experience; specifically, there is a lack of such knowledge that is grounded in existing theory on user experience. In this paper, we contribute to establishing this. We present the findings from a survey study where more than 200 chatbot users reported in their own words on positive and negative experiences with chatbots. This approach allowed the participants to highlight the attributes of their chatbot user experience that were most salient to them. As such, the identified attributes should be of particular relevance to chatbot research and development.

The paper is structured as follows. First, we provide a brief overview of relevant background before detailing the research problem and method. Then, the findings are presented, with particular attention to the participants' reports of good and poor experiences. Finally, we discuss the findings, provide guidelines for chatbot development, and suggest relevant future research.

## Background

### User experience

In the general research literature on interactive systems, it is well established that user experience is decisive for the uptake of such systems in user populations [28]. Researcher and practitioner interest in user experience is a consequence of acknowledging that interactive systems may also cater to users' desires for outcomes which do not directly concern productivity—such as using interactive systems for engaging or immersive activities. User experience has been studied for a range of application domains, from games [2] and websites [45] to mobile phones [33] and larger hardware systems such as ATM terminals [44].

In this paper, we position our understanding of user experience within what Forlizzi and Battarbee [18] would refer to as *user-centred models*, as we mainly consider the perspective of people who use chatbots. Following the international standard for human-centred design of interactive systems, we understand *user experience* as a 'person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service' [29, p. 3]. In line with Law et al. [35], we consider user experience to be dynamic, context-dependent and subjective.

The dynamic character of user experience implies that the experience of a particular interactive system may change over time as users grow accustomed to the system, run in to

difficulties, or simply lose interest [33]. Thus, it is critical that interactive systems regularly and consistently provide users with useful or pleasurable episodes of use.

The context-dependent character of user experience implies that users' perceptions of their system interaction depend on factors that are partly beyond the control of the system developer. User experience depends on the time, place and purpose [24].

The subjective character of user experience implies that it is not immediately observable to developers and service providers. As such, user experience represents a methodological challenge of gaining access to users' internal perceptions, emotions and reflections [25].

User experience can be investigated from different perspectives. Law et al. [34] distinguished between design-based and model-based perspectives, where the latter concern the representation of user experience through established constructs to facilitate comparison and generalization. Within the model-based perspective, a range of frameworks or models have been proposed addressing experiential attributes, such as enjoyment, beauty, satisfaction, flow and trust, as well as productivity-oriented attributes, such as usefulness and ease of use [28]. The proposed attributes of user experience may typically be grouped as pragmatic or productivity-oriented on the one hand and hedonic or engagement-oriented on the other. To support an initial understanding of the chatbot user experience, a framework facilitating an exploration of these two broad groups of attributes will be of particular interest.

Hassenzahl's framework for analysing user experience [23] has been formative for much of the user experience literature. The framework specifically addresses the distinction between *pragmatic* and *hedonic* attributes of an interactive system. These attributes summarize the intended (by the designer) or perceived (by the user) characteristics of an interactive system. According to Hassenzahl, pragmatic attributes concern instrumental characteristics, such as usefulness and usability; that is, they relate to whether an interactive system provides relevant task-oriented functionality in an accessible and easy-to-use manner. Hedonic attributes involve characteristics of relevance for the mental or emotional well-being of the user, that is, emotionally salient or rewarding aspects of the interactive system. Hedonic attributes may, for example, concern stimulating aspects of the interactive system, its ability to communicate identity, or its evocative character.

Interactive systems vary in the degree to which they are characterized by pragmatic and hedonic attributes. Systems with strong pragmatic and weak hedonic attributes are beneficial for instrumental purposes, that is, to achieve behavioural goals. Systems with strong hedonic and weak pragmatic attributes are seen as self-oriented, beneficial for stimulating, evocative or identity forming purposes.

Hassenzahl [23] argued that the combination of strong pragmatic and hedonic attributes is desirable and should be seen as an ultimate design goal. However, he also noted that there likely will be an imbalance between the two attribute groups, leading to products that are either more pragmatic or more hedonic.

Hassenzahl's framework has been applied in the study of a wide range of interactive systems, including consumer products [12] and mobile applications [3]. The framework has also been used as a basis for investigating other aspects of user experience such as goodness and beauty [45] and as a basis for the AttrakDiff measurement instrument for user experience [26].

In our study, we have chosen the pragmatic-hedonic framework of Hassenzahl as a basis for our analysis because it serves to shed light on the range of different perceptions and experiences users report from their interactions with chatbots.

### Current chatbot user insight

We understand *chatbots* as machine agents with which users interact through natural language dialogue, by text or voice [4]. This definition is in line with how the term is used in media and the industry and corresponds to what members of the natural language processing (NLP) community refer to as *dialogue systems* or *conversational agents* [30]. Our inclusive use of the term *chatbot*, which may be broader than definitions used in the NLP community, reflects a view that it has become increasingly difficult to discriminate between the traditional categories of systems for natural language dialogue—task-oriented versus social or text-based vs. voice-based—because the same applications often incorporate elements of several such categories. Chatbots are typically provided as a means to address one or more specific user goals [41]; hence, the vast majority of operational chatbots are likely to have a task-oriented characteristic. However, exceptions exist—most noteworthy, chatbots designed for social interaction and relationships—such as Mitsuku[1] (social chatter) and Replika[2] (social relationships).

The recent burst of interest in research on chatbots and conversational systems has led to a substantial increase in the body of research pertaining to chatbot use. In particular, there has been considerable research on users' behaviour with chatbots, user perceptions and preferences regarding chatbots and chatbot use in particular contexts.

To address chatbot user behaviour, researchers have studied how users interact with chatbots in both laboratory and real-world settings. For example, Hill et al. [27] compared user interactions with a chatbot to user interactions during chats with other users and found that chatting with chatbots involved longer conversations with more and shorter messages and less richness than what was found in chats with other users. Porcheron et al. [39] investigated real-world use of a voice-based conversational agent and revealed that the real-world context introduces challenges to the conversation and leads to frequent breakdowns in the chatbot interaction. Studies such as these provide insights into how users behave as part of, or in response to, their interactions with chatbots. As such they provide valuable guidance to chatbot interaction design, for example to help mitigate usability problems and strengthen conversational repair.

In addition to research on chatbot usage, there is also growing interest in investigating user perceptions and preferences regarding chatbots. For example, Brandtzaeg and Følstad [5] investigated users' motivations for chatbot use and found that productivity was the main driver for use, followed by entertainment, social purposes and an interest in chatbots as a novel technology or user interface. Thies et al. [43] investigated user preferences for different chatbot personas and found that personas reflecting both productivity and engagement were preferable to users. Luger and Sellen [37] investigated user experiences with voice-based chatbots and found that it was challenging for users to properly understand chatbot capabilities and to interact efficiently with chatbots; they suggested that play might be a useful entry-point for learning about chatbot features and efficient interactions.

Relevant insights about chatbot use and user experience can also be drawn from the dialogue systems evaluation literature, for example in terms of frameworks and metrics for dialogue evaluation [e.g. 36, 46] that target constructs such as effectiveness, efficiency and user satisfaction for task-oriented systems, and appropriateness and human likeness for systems oriented towards small talk [11]. Also, dialogue system challenges, such as the Alexa Prize [31], where teams compete to develop voice-based chatbots for engaging social interactions, contribute substantially to increasing our understanding of user perceptions and preferences regarding chatbots. However, what is lacking in the current research literature is insight into how pragmatic and hedonic attributes, separately or in combination, contribute to the overall chatbot user experience.

Chatbot use and user experiences have been studied in a wide range of contexts, including education [19], mental health [16] and information services [8]. In such studies, users' subjective perceptions of chatbots have been seen as relevant assessment criteria. For example, Fitzpatrick et al. [16] investigated users' best and worst experiences during their use of a therapeutic chatbot. Crutzen et al. [8] asked users of an information chatbot for youth about whether they

---

preferred to seek information from the chatbot or from other sources. What is currently lacking is research on chatbot user experience that spans a broader set of contexts to gain insight into aspects of chatbot user experience that are relatively stable across contexts. Interestingly, a good deal of the research on chatbot user experience is conducted as Wizard-of-Oz studies; that is, the participants do not interact with an actual chatbot but with a human simulating a chatbot [e.g. 1, 6, 43], which suggests the challenges of investigating user experience for chatbot concepts in early stages of development. This, in turn, highlights the importance of conducting research on user experience for chatbots that have been made available to users.

## 3 Research question

The body of research on user experience for interactive systems in general, as well as the emerging literature on chatbot user insights, provides a good starting point for understanding how chatbot user experience is formed and the effect it has on user behaviour.

However, there is a lack of studies that systematically address chatbot user experience with regard to pragmatic and hedonic attributes. This gap in current knowledge limits researcher and practitioner understanding of chatbot user experience in terms of the breadth of the user experience construct. Hence, we find it valuable to conduct a study of chatbot user experience that reflects a pragmatic-hedonic framework.

Furthermore, there is a gap in the literature concerning explorations of chatbot user experience across domains and chatbot implementations. Specifically, we find it important to investigate chatbot user experience across a range of chatbots and domains in order to gain an overview of chatbot characteristics and use that can be seen as particularly important regardless of domain. At the same time, such an investigation would need to acknowledge the substantial individual variations in chatbot user experience.

On this basis, we established the following exploratory research question for the study:

> What are the key characteristics of chatbot user experience?

The exploratory character of the research question invites an investigation that considers both the pragmatic and hedonic attributes of chatbots and also attributes that fall outside this conceptual dichotomy. Specifically, the research question motivates investigations of factors that drive good and poor user experiences. Such factors could be contextual, could pertain to the design or implementation of the chatbot or could be a consequence of user characteristics. However, to be of relevance for this study, these factors should be recognised by the user as actually driving or constituting the good or poor user experience.

During our explorations, we were intrigued to see that some respondents reported on the use of chatbots by children or youth. We therefore decided to also include an investigation of the effect of age with respect to drivers of chatbot user experience.

## Method

To gain insight into chatbot user experience, we saw it as necessary to gather data from a sample of participants experienced in chatbot use. Furthermore, because we aimed to explore characteristics of chatbot user experience across chatbots and application domains, we needed a sample that included participants who were experienced with different chatbots.

For this purpose, we applied a questionnaire study. Here, we aimed to involve a relatively large sample of participants with diverse backgrounds in chatbot use and have them report on their chatbot user experience.

### Participant recruitment

Chatbots are not yet in common use for the majority of information and communication technology users. Hence, participant recruitment and filtering were important parts of the study method. We decided to recruit a US sample because important platforms for chatbot use, such as Facebook Messenger, Skype, and Kik, target this geographical market in particular. We also decided to involve only relatively young participants (16–55 years of age), as we saw these as more likely to be chatbot adopters given their more frequent use of messaging platforms.

Data were collected in two waves: April and June 2017. The recruitment of study participants and administration of data collection were done by Survata, an independent research company. Survata samples participants through partnerships with online publishers. Participants gain access to premium content as an incentive for participation. In our study, participants were invited to complete the questionnaire following an introductory filtering question about their experience with chatbots. Findings based on the first wave data collection have previously been presented by Brandtzaeg and Følstad [5].

### Study material

The questionnaire consisted of 17 questions, including two items of particular interest for this study. These were open-ended questions where the participants responded in free text, using their own words. The other questions concerned

participant demographics (gender and age), aspects of chatbot use (e.g. platforms for chatbot use, frequency of chatbot use, duration of chatbot use, characteristics of used chatbots and motivation for chatbot use).

The first of the two items of particular interest for this study concerned positive episodes of chatbot use. The participants were asked to 'tell about one really good experience that you have had with a chatbot'. They were specifically requested to use their own words and to be as detailed as possible.

The second of these two items concerned negative episodes. Here, the question was posed in two different ways. For the first wave of data collection (April 2017), the participants were asked to 'tell about a really bad experience you have had with chatbots' and were asked to use their own words and provide as much detail as possible. For the second wave (June 2017), the participants were asked, 'Have you ever stopped using a chatbot? Please explain the most important reason to stop using it.' The reason for modifying how this question was asked across the two waves of data collection was that a surprisingly large proportion of participants in the first wave did not report negative episodes of use. We therefore decided to change the question to what we saw as a relatively stronger negative wording. The findings for these two versions of the negatively worded questionnaire item, however, did not differ substantially and were therefore analysed as one data set.

The design of the two qualitative questionnaire items was motivated by the critical incident technique (CIT), a qualitative approach to data collection with roots in the field of psychology [17]. Using CIT, researchers gather users' stories about salient incidents during their exposure to a particular phenomenon in order to identify and understand the factors that are important in driving their experience. This technique is particularly beneficial because it provides a way to gather rich insight into those incidents or characteristics that are most important in determining user experience from the perspective of the user [21]. As such, CIT is a potentially valuable technique for investigating chatbot user experience. We are not aware of other studies that have applied a CIT approach for this purpose.

## Data analysis

Prior to analysis, the dataset was filtered to ensure that all participants had sufficient experience with chatbots and that there was sufficient variation in the chatbots used by the participants. This filtering was based on the participants' responses to the question, 'What is the name of the latest chatbot you used?'. We saw this question as a good indicator that (a) the participants actually had knowledge of chatbots, and (b) the participants were not confusing the term chatbot with chat platforms for human–human interaction.

All participants who reported a name corresponding to a known chatbot or who were able to describe such a chatbot if they didn't remember the specific name were included for analysis. The responses of the remaining participants were filtered out.

The participant reports were analysed through thematic analysis [14], where a set of codes was identified on the basis of the themes emerging in the qualitative data. This set of codes was then grouped according to Hassenzahl's [23] framework of pragmatic and hedonic attributes and used to analyse all participant reports. Note that some of the codes represent attributes not anticipated in the pragmatic-hedonic framework, for example attributes that reflect the social and human-like characteristics of chatbot interaction.

Each participant report could be coded as corresponding to one, several or none of the codes. Participant reports that did not correspond to any of the codes were coded as 'other'.

Coding participant responses to reflect pragmatic or hedonic attributes also allowed us to investigate the effect of age on the likelihood that one or the other of these two attribute types would be emphasized in participant responses. To determine this, we conducted two bivariate correlation analyses: one for participants' tendency to report on pragmatic attributes and one for their tendency to report on hedonic attributes. These two analyses were conducted using the statistical software package IBM SPSS Statistics 25.

## Results

### About the participant sample

The sample consisted of responses from 207 valid participants. Of these, 62% were male and 38% were female. Mean age was 27 years (SD = 8). The participants used chatbots on a variety of messaging platforms, notably Facebook Messenger (79%), Skype (54%), Kik (38%), Viber (12%), Slack (10%) and Telegram (4%). Further, 65% reported using chatbots daily or weekly, 48% reported having used chatbots for 3 or more years and 40% had experience with Google Assistant.

The chatbots reported as the most recently used by the participants reflected a wide range of chatbots: for productivity purposes, marketing, customer service and entertainment. The most reported recently used chatbots were virtual assistants such as Google Assistant (18%), Siri (7%), Alexa (4%) and Cortana (4%) and chatbots for social chatter such as Cleverbot (11%), Eviebot (3%), Mitsuku (1%), SimSimi (1%), Zo (1%) and the no longer operative Smarterchild (3%). The frequent mention of Google Assistant as a recently used chatbot may be due to its availability on the Android operating system.

**Table 1** Coding categories for positive chatbot user experience reports, with associated attribute type, descriptions and frequencies ($N = 207$)

| Type | Category | Description | Freq (%) |
|---|---|---|---|
| Pragmatic | Help and assistance | The chatbot is reported to provide customer support or training, personal assistance, or help with a particular task at hand. Efficiency and ease of access were often highlighted | 34 |
| | Information and updates | The chatbot is reported to provide updates and general information—often sought on a routine basis—such as news, weather and online searches | 8 |
| Hedonic | Entertainment | The chatbot interaction is described in words reflecting engagement and enjoyment | 29 |
| | Novelty and inspiration | The novelty of chatbots, or the inspirational value of the interaction, is accentuated | 8 |
| Other | Social | The positive chatbot user experience is reported to concern social interaction, either human–human or human-chatbot | 7 |
| | Humanlike | The humanlike character of the chatbot is accentuated as a source of a positive user experience | 4 |
| | Miscellaneous | Less frequent coding categories, including users reporting to challenge the chatbot, chatbots for therapy use and chatbots used for surveys | 9 |

## Positive chatbot experiences

Participants reported a broad set of positive chatbot experiences. The reports were on average 105 characters long (SD = 90). Nearly all reports provided sufficient detail to identify one or more characteristics of the experience.

Although we asked the participants to report on a specific positive episode, not all of them did. Rather, 45% reported on a specific episode, whereas 22% reported on their overall experience with a specific chatbot. The remaining participants made high-level reports of chatbot attributes regarded as positive without mentioning a specific episode or chatbot. To provide a feel for what the participant reports look like, we include the following two quotes, which exemplify reports of a specific episode (P37) and reports of a specific chatbot (P40).

> I actually recently interacted with a chat bot about a complaint I had with a company. The chat bot informed me the correct way and persons to send my information to. It was quick and easy and I really appreciated this since I was already quite annoyed (P37)
>
> I get Cortana to tell me a joke—on Windows 10 (P40)

Following Hassenzahl's [23] pragmatic-hedonic framework for user experience, the participant reports were analysed with regard to whether they reflected pragmatic attributes of an interactive system, such as usefulness and usability, and/or hedonic attributes, such as facilitating evocative or stimulating experiences.

In the user reports, we found an appreciation for both pragmatic attributes and hedonic attributes in participants' detailing of positive chatbot user experiences; 42% of the participant reports reflected pragmatic chatbot attributes and 36% highlighted hedonic attributes. In addition, 20% of the reports reflected codes that are not directly related to pragmatic or hedonic attributes. The most frequent of these additional codes concerned the social aspects of an interaction (7%) and the chatbot's character as humanlike (4%).

The distribution of codes for the participant reports of positive chatbot user experiences is provided in Table 1. Details concerning each coding category are provided following the table.

### Pragmatic attribute: help and assistance (34%)

The participant reports strongly reflected the importance of perceived usefulness or practical value for positive user experience. When asked to report a particularly good episode, participants often reported on getting assistance or help from the chatbot. A number of the reported episodes concerned customer service support and also instances of training or coaching through the chatbot. Other episodes concerned personal assistance, such as setting reminders for tasks or getting help with a specific task at hand, as in this quote:

> I asked what good places there was around me to eat and it brought up a list and i chose from it. Now the place is one of my favorite places to eat at (P21)

The instrumental or pragmatic characteristics of chatbots were clearly apparent in the reported episodes, where task achievement and efficiency in particular were highlighted as important in numerous participant reports. Participants reported receiving help in situations where they were pressed for time because of an urgent problem or a circumstance where they needed information quickly. Participants also made particular note that the assistance was efficient and easily accessible, as in the following example:

> The chatbot for customer support for my wireless carrier was a great help! I didn't have to wait on a representative to become available, I was able to find out what I needed to know about different plans and their pricing. So much better than sitting on hold or waiting an hour for someone to message me back (P23)

### Pragmatic attribute: information and updates (8%)

While help and assistance for a particular task was by far the most frequently reported category of positive user experience episodes, some participants instead reported on the pragmatic benefit of chatbots for general information searches or more routine updates, such as news reports and weather forecasts. These participants highlighted the chatbots' support for retrieving general online information or daily updates, rather than getting help in a particular situation. Participants reported gathering information through a chatbot that they would otherwise be able to access through a search engine. Participants also reported using chatbots for doing their everyday checks of information important to their daily routines, for example in the following quote:

> I use google assistant to do simple tasks on my phone every morning when I am waking up and need to know the time and weather [...] (P59)

### Hedonic attribute: entertainment (29%)

In their reports, participants indicated substantial appreciation for hedonic chatbot attributes. When reporting on such non-pragmatic aspects of chatbot user experience, participants typically highlighted the entertainment value of chatbots. Entertaining chatbot episodes were presented in ways that indicated they were seen as stimulating and contributed to the participant feeling happy and engaged. Participants used words such as 'fun', 'entertaining' and 'cool', as in the following example:

> It was funny. I asked it if it liked me and it asked me if I like me (P207)

Participants who reported the entertainment value of chatbots typically referred to situations where they engaged in small talk with a chatbot. That is, they often did not have a particular task to be resolved but rather saw the chatbot as a means of involving themselves in a pleasing activity. Specifically, they reported that the chatbots' ability to be funny and witty was a source of pleasure, or they reported that the chatbot was something they could joke with or turn to when bored. An example of such use is reflected in the following quote:

> Chatbot and I just kept talking random things, that when looked at after made some sense. it was fun (P128)

To our surprise, quite a few participants who reflected on hedonic aspects of the user experience reported on the use of chatbots by children. Some of these reports were from their own childhood; for example, participants reported that conversations with chatbots as a child were a source of entertainment in the company of their friends, or a source of relief as a teenager when they were bored. Other participants described experiences as parents—observing their own children engaging with chatbots, either on the initiative of the parent or through the child's own initiative. The finding that chatbots serve as a source of stimulation and engagement for children is interesting, as it suggests the potential of chatbots to stimulate playful social interaction for and possibly also among children.

> My earliest memories of artificial intelligence are with an online chatbot called SmarterChild. I remember it being pretty funny sometimes, witty and intelligent, almost like it was a real person behind the character typing his responses (P132)

### Hedonic attribute: inspiration and novelty (8%)

Some of the participants who highlighted hedonic chatbot attributes in their descriptions of good chatbot experiences reported an inspirational episode or a general sense of novelty in chatbots.

Among the participants who reported on the chatbot as inspirational, some described the episode as 'eye-opening' or described how they were able to talk to the chatbot about a topic that engaged them, such as pets or food. Such reports in part reflect the potentially evocative character of chatbots and in part reflect the potential of chatbots to adapt to topics with which the user identifies, as in the following participant quote:

> I had a pleasant conversation about my life with a chatbot. I talked about my family and my feelings (P182)

Some participants also reported being excited or engaged by their perception of chatbots as a novel and fascinating way to interact with computers. In these reports, some participants explained how they saw it as amazing to actually have a conversation with a computer, and some also described how they had tried to test the degree to which the chatbot is able to act like a human. The following quote exemplifies participant reports belonging to this coding category:

> I had an interesting experience trying out artificial intelligence through small talk with a chatbot. You could tell it wasn't human but it was interesting nonetheless (P83)

### Other attributes: social and humanlike (11%)

While the coding categories for pragmatic and hedonic attributes are reflected in most of the participants' reports, some reports referred to attributes that did not readily fit into the pragmatic-hedonic framework but were still relevant to

**Table 2** Coding categories for negative chatbot user experience reports, with associated attribute type, descriptions and frequencies (*N* = 207)

| Type | Category | Description | Freq (%) |
|---|---|---|---|
| Pragmatic | Interpretation issues | The chatbot is reported to misinterpret requests or input, or to provide an answer that does not fit with the question | 11 |
| | Unable to help | The chatbot is reported to be unable to assist the participant in solving a particular task or to be unable to provide help in general | 11 |
| | Repetitiveness | The chatbot is reported to ask the same questions or provide the same line of answers repeatedly, which is experienced as obstructing the user from getting help or assistance | 4 |
| Hedonic | Strange or rude responses | The chatbot is reported to give responses that are improper or embarrassing | 7 |
| | Unwanted events | The chatbot is reported to be the source of unwanted contact, actions or content | 6 |
| | Boring | The chatbot interaction is reported to be boring—either immediately or after a period of use | 4 |

user experience. In particular, this was the case for participant reports about the social value they received from using the chatbot (7%) as well as the perceived experiential benefit of the chatbot being humanlike (4%).

Social value typically involved enjoying a social situation with the chatbot. In these reports, the participants described how they appreciated the social interaction with the chatbot. That is, even though they were aware that the chatbot is a machine, the social interaction was seen to hold value in itself, as in the following example:

> Chatbot helps me get my day moving when I don't talk to anyone (P141)

For some participants, the chatbot was used to support social interaction with other (human) users, as for example in group chats. Here, the chatbot could serve instrumental purposes in a social interaction, such as providing linked content on topics of conversation or helping to get conversations or groups started.

> I used chatbots to send links to websites mentioned in Skype conversation. It was very convenient way to make the conversation more efficient (P45)

The identified humanlike attribute was in many ways associated with the social attribute. Here, participant reports explained how chatbot characteristics that are almost human may contribute positively to user experience. The humanlike character of chatbots was noted by some participants, but fewer than might have been expected given that this is often seen as a prominent chatbot attribute.

> I use to ask them all kinds of questions till they had a whole conversation with me told me where they were from an how they worked as a waitress at a bar it was such a funny conversation i actually thought the chatbot was a real person (P34)

In reports that emphasize the social or humanlike attributes of chatbots, these attributes were often discussed together with hedonic or pragmatic attributes. We nevertheless found it important to single out the social aspects of chatbots as

reflecting a distinct attribute outside the group of hedonic attributes because the pragmatic-hedonic framework does not specifically address social interactions with interactive systems. It should be kept in mind that our coding scheme allowed any user report to have multiple codes associated with it, so this represented no challenge in terms of coding.

## Negative chatbot experiences

We asked the participants to report on poor or unpleasant chatbot user experiences. In these reports, the most frequent characterizations of the chatbot and the user experience involved pragmatic (23%) and hedonic (16%) attributes, with pragmatic attributes being the most prevalent.

The distribution of codes for the participant reports on negative chatbot user experiences is provided in Table 2. Details concerning each coding category are provided following the table.

Before we present the findings, it should be noted that the participants were much less inclined to report negative experiences than they were to report positive ones. In fact, 41% of the participants reported that they had not had a bad experience with a chatbot (first wave of data collection) or they had not stopped using chatbots (second wave of data collection). It should be noted that no participants skipped this question (mandatory question); rather, the participants who reported not having negative experiences did so in their own words.

This lack of negative episodes was a surprise to us as chatbots are still a relatively immature type of interactive system. One possible explanation for this lower frequency of reports of negative user experiences is that the respondents are relatively early adopters [40] and are hence more tolerant of technical issues or interaction breakdowns in this emerging technology. Another explanation may be a participant response bias, where study participants might hesitate to make negative assessments of a technology under study [10]. It could also be that different phrasings of our request for negative experiences, for example asking for negative

experiences in general, would have increased the frequency of negative experience reports.

Nevertheless, the number of reports actually detailing negative user experiences was more than sufficient to establish coding categories and provide insight into chatbot attributes that potentially drive such user experiences.

### Pragmatic attribute: interpretation issues (11%)

Not surprisingly, the chatbot attributes reported to drive poor user experiences often concerned pragmatic attributes—specifically, usability. In particular, interpretation issues were prominent. A substantial number of the participant reports of negative episodes concerned the perceived challenge of making the chatbot understand what the user was trying to tell it.

Interpretation issues could be framed as a general frustration that chatbots sometimes need questions or requests to be repeated, or that they are not able to correctly interpret the user input at all. For example, participants detailed how they had to adapt their way of expressing themselves in order to be understood by the chatbot, or they reported that they stopped using chatbots that were unable to understand their input.

> Each time I answered a question it would not read my response or respond asking me to repeat my answer. I eventually got annoyed and exited out of it (P43)

### Pragmatic attribute: unable to help (11%)

Linked to the issue of interpretation, poor chatbot user experience was often reported when the chatbot was unable to help. In usability terms, participants reported on low effectiveness in the chatbot. Such low effectiveness may be deeply frustrating, as it may completely compromise the potential benefit of the chatbot. The reported lack of help could be due to the chatbot providing an answer that was too generic, and in some cases due to the chatbot's answer being irrelevant to the participant's task at hand. Some participants also noted that they did not trust chatbots or suspected that the chatbot in question provided false information. The following participant quote exemplifies the experience of users when chatbots are unable to help:

> When I ask a question, some of the info they provide are completely irrelevant. Still kinks to be ironed out. It's still easier to just Google search for answers to your questions (P68)

### Pragmatic attribute: repetitiveness (4%)

A third pragmatic chatbot attribute reported to generate negative user experience was the repetitiveness found in some chatbots; that is, the experience that the chatbot just keeps reiterating the same questions or responses. Participants reporting on repetitiveness described chatbots that ask the same thing over and over without making progress towards the intended task goal. This may be in the context of customer support or information seeking, for example where the chatbot is not able to derive the information needed from the participant in order to progress in its routine. Some participants also reported on chatbots that repeatedly suggested taking the same steps to resolve a problem, even though those steps have already proven unfruitful.

> There have been many occasions where a bot has either looped around with its inquiries, and/or not had the info I needed. It made me feel like I had just wasted a good deal of time (P14)

### Hedonic attribute: strange or rude responses (7%)

Whereas pragmatic aspects dominated the reports of negative episodes, some reports reflected the hedonic or emotionally charged aspects of the user experience—in particular with regard to a tendency for chatbots to provide strange or rude responses. Rude responses are understood as chatbot responses that are at odds with what the participant sees as acceptable; strange responses are understood as chatbot responses that are seen as off-topic—though not in a task-oriented context. Some participants reported that rude and strange responses were embarrassing, possibly because they were seen as breaking with the promise of intelligence that natural language interaction suggests.

> When i asked a question at a friends house and it brought up an answer that was not only irrelevant but dirty (P66)

In a sense, strange or rude responses from chatbots may be seen as somehow related to chatbots being witty and funny. Different users may even interpret the same chatbot response as either funny or embarrassing.

### Hedonic attribute: unwanted events (6%)

In addition to strange or rude responses, the participant reports contained references to unwanted events as a negative hedonic chatbot attribute. Participants reported on chatbots that had contacted them at times or in ways that were unsuitable, chatbots taking actions they did not want them to take or chatbots presenting unwanted content. In particular, participants reported on chatbots initiating contact at times when the participant would rather not be disturbed, as in the following example:

I was sleeping one time and the chatbot started texting via messenger and i got really angry because i was sleeping real good and it woke me up (P21).

Such unwanted contact was typically reported to be disturbing, likely because it was experienced as the chatbot invading the user's private life or as the user losing control of how and when the chatbot interacts with them.

Likewise, unwanted actions from chatbots were seen as frustrating and could potentially hold substantial negative consequences:

I used the chatbot to send an email and it sent it to the wrong person in my contacts (P18)

### Hedonic attribute: boring (4%)

In addition to the two hedonic chatbot attributes discussed above, some participants noted that their negative chatbot user experience was due to the perception of the chatbot as boring. Chatbots are an emerging technology, so it is not surprising that a chatbot may be perceived as boring once the novelty wears off. As such, the perceived boring character of some chatbots can be seen as a likely outcome, considering that the novelty of chatbots was found to be a driver of positive user experiences for some participants.

In particular, boredom or lack of experiential value over time were seen as important reasons to stop using chatbots, as seen in the following participant quote:

I stopped using MurphyBot because the novelty of it wore off. Since it mostly was used to search for and meld images together, it got boring pretty quickly (P79)

### Age differences

Because participants often mentioned childhood events when sharing positive episodes of chatbot use, we wanted to investigate whether age was related to which chatbot attributes the participants highlighted in their reports.

We had not initially planned this investigation. However, reports of childhood memories or experiences of enjoying talks with chatbots suggested the relevance of doing such an investigation. For this purpose, we conducted two bivariate correlation analyses to investigate the degree to which age predicts whether a participant will report pragmatic or hedonic chatbot attributes respectively when describing positive user experiences. We applied the Spearman rank correlation, as the age data was found to be non-normal following a Shapiro–Wilk test of normality.

In these analyses, we found that greater age was positively associated with the tendency to report on pragmatic chatbot attributes, that is, *help and assistance* and *information and updates* (Spearman's $Rho = 0.31$, $p < 0.001$). In contrast, greater age was negatively associated with the inclination to report on hedonic chatbot attributes, that is, *entertainment* and *novelty and inspiration* (Spearman's $Rho = -0.25$, $p < 0.001$).

In other words, older participants tended to report on pragmatic chatbot attributes, whereas younger participants tended to highlight hedonic chatbot attributes. Effect sizes were medium following Cohen's rules of thumb [15].

## Discussion

In the discussion section, we first summarize and reflect on our findings regarding the research question on the key characteristics of chatbot user experience. We then provide four high-level lessons learnt that may benefit future chatbot development. Finally, we discuss the study limitations and point out relevant future work.

### Characteristics of good chatbot user experience

#### The potential complement of pragmatic and hedonic attributes in chatbot user experience

Theories of user experience with interactive systems suggest that pragmatic and hedonic aspects are critical [35]. This seems to also hold true for chatbot user experience. In particular, pragmatic aspects of user experience are important. Good chatbot user experience is created through useful and efficient interactions. At the same time, users also acknowledge the importance of hedonic aspects of user experience. Specifically, users emphasize the entertainment value of chatbots, although the potentially inspirational value of chatbots and the novelty of chatbot interaction are also mentioned by some as potentially beneficial to chatbot user experience. The potential benefit of combining strong pragmatic and hedonic attributes in chatbots is not surprising seen from the perspective of Hassenzahl's framework [23]—as designs that are strong in both attribute groups are seen as desirable.

These findings echo previous work on user motivations and preferences regarding chatbot use. Brandtzaeg and Følstad [3], in their study of chatbot user motivations, found that productivity was the key motivator for most users, followed by entertainment. Thies et al. [43] found that potential chatbot users prefer a chatbot personality that reflects both productivity and engagement. For chatbot service providers, it may thus be beneficial to design for both pragmatic and hedonic chatbot attributes, provided the application area is appropriate for this.

### The relative importance of pragmatic and hedonic attributes with regard to chatbot type

While it may be beneficial to complement pragmatic attributes with hedonic attributes in chatbot user experience, it is important to do so in consideration of chatbot type. A broad distinction has traditionally been made between task-oriented chatbots and chatbots oriented towards social interaction [7].

For task-oriented chatbots, through which users aim to achieve specific goals, our findings suggest the importance of leveraging hedonic attributes to complement a user experience that is otherwise merely a pragmatic interaction. For example, chatbots for education [19], health [16] and customer service [38] may provide a more compelling user experience if the hedonic attributes of the interaction are also addressed in the chatbot interaction. For some application areas, however, such as areas involving service provision associated with high perceived risk or importance, a more pragmatic orientation may be required. Conversely, for application areas characterized by easy entertainment, a more hedonic orientation may be seen as valuable. The hedonic attributes of user experience may also be leveraged to a greater or lesser extent depending on the users' experience with the interactive system. For example, Luger and Sellen [37] noted that virtual assistants make strategic use of entertainment-oriented and playful interactions to onboard users and help them get used to interacting in a conversational format.

For chatbots oriented towards social interaction [42], pragmatic attributes of user experience are typically less emphasized. In this case, concern will instead be oriented towards providing a compelling and engaging experience and, consequently, our findings concerning the benefit of pragmatic attributes of user experience may not be as relevant here. Nevertheless, we believe that designers and developers of chatbots oriented towards social interaction may also benefit from leveraging both hedonic and pragmatic attributes of user experiences. Examples of this may be that chatbots oriented towards social interaction may provide an improved user experience by also being able to help with specific tasks, such as setting reminders or looking up information. For example, in the Alexa Prize challenge, where dialogue system research teams compete to develop engaging conversational interactions through Amazon Alexa, facilitating knowledge-rich conversations is seen as important for sustained user engagement [31]. Socially oriented chatbots may also to a greater degree leverage a variety of hedonic attributes, such as offering more inspirational content.

### Characteristics of poor chatbot user experience

#### Breakdowns may be due to both pragmatic and hedonic attributes

Our findings show that pragmatic chatbot attributes may be the cause also of detrimental user experiences, specifically when the chatbot is challenged to correctly interpret the users' intention or is unable to provide effective assistance. In such cases, user experience suffers. As such, while pragmatic attributes may be key to providing a good chatbot user experience, a breakdown in the pragmatic application of the chatbot will lead to poor user experiences. These findings suggest that chatbot service providers should carefully consider the possibly negative implications of launching beta versions of chatbots before the chatbots can correctly interpret the vast majority of user requests and offer the help users need. Premature launch may be detrimental to user experience, in particular for chatbots that are intended to serve a pragmatic purpose, such as customer service or providing information. As suggested by the work of Porcheron et al. [39], such breakdowns in pragmatic application may also be a consequence of the context of use; for example, a chatbot applied in the context of a group of family members may perform differently than a chatbot applied by a single user.

However, chatbot user experience may also suffer from negative hedonic attributes. In particular, when the chatbot is found to break the contract suggested by its natural language interaction and offers out-of-place or even rude responses, users may experience strong negative emotions in response. This finding points to an important challenge for chatbot service providers. As the context of chatbot use is difficult to predict, chatbot responses that are well intended by the chatbot content provider may seem inappropriate in the context of use. Hence, great care must be taken to avoid chatbot content that can be easily misinterpreted or that can hold highly different connotations in different contexts of use.

#### Chatbot type and the importance of breakdowns due to pragmatic and hedonic attributes

Chatbot type may be important to the relative importance of pragmatic and hedonic attributes for user experience. Task-oriented chatbots in particular may be prone to breakdowns concerning pragmatic attributes. If the chatbot fails to interpret users' requests or is unable to provide the needed help, user experience for the task-oriented chatbot will clearly suffer.

Breakdowns in terms of interpretation issues and failure to provide requested help may be less of an issue in chatbots oriented towards social interaction. Rather, in such chatbots, user input that the chatbot is unable to interpret correctly

may be mitigated by the chatbot diverting the conversation to a different topic or using other socially acceptable mechanisms for conversational repair to avoid its interpretation failure to be detrimental to user experience.

Task-oriented and socially oriented chatbots also differ in terms of the potential risk involved in leveraging hedonic attributes of user experience. For task-oriented chatbots, leveraging hedonic attributes could lead the user to see the chatbot as more humanlike, in turn paving the way for erroneous assumptions regarding the chatbot's capabilities [39]. Hedonic chatbot attributes in task-oriented chatbots thus need to be designed with great care to provide the intended positive user experience. This potential risk may be less important in chatbots oriented towards social interaction, as such interactions may be less dependent on the user having an adequate model of the capabilities of the chatbot.

## Age and variations in attributes of user experience

It is interesting to note that chatbot user experience seems to differ with age. The variations in user experience across user groups may be expected given its subjective character and potential for individual variation [25]. Specifically, younger users may be particularly sensitive to playful and emotionally engaging chatbots, whereas older users may be preoccupied with the efficiency and effectiveness of chatbots.

Such variations across age groups may also indicate which age groups are more inclined to benefit from different types of chatbots. In particular, chatbots for social interaction may potentially be seen as more valuable by younger user groups, provided that pragmatic attributes are relatively less present in such chatbots. Task-oriented chatbots, however, could be tailored to different age groups by varying the balance between pragmatic and hedonic attributes.

## Lessons learnt

The main contribution of this paper is the findings already presented and discussed. Drawing on this contribution, we now formulate four high-level lessons learnt—for the benefit of future chatbot development. While in the above "Characteristics of poor chatbot user experience" and "Age and variations in attributes of user experience" sections we discussed the particular contributions of this study, here we summarize lessons that are suggested by the findings. The lessons learnt address task-oriented chatbots in particular.

1. *In task-oriented chatbots, usefulness is king* For task-oriented chatbot applications, solving users' problems and helping users reach their goals in an effective and efficient manner are key to providing good chatbot user experiences. For sustained interest, it is important to provide valued help and assistance, and for practically all chatbot applications, it is critical to correctly interpret the user's intentions and provide adequate responses. Even though chatbots are still an emerging type of interactive system, it is important for service providers to take great care that their chatbots serve their intended purposes and that these purposes are valued as useful by their users.

2. *Hedonic attributes may strengthen user experience in task-oriented chatbots* For many task-oriented chatbot applications, user experience can be strengthened by blending pragmatic and hedonic chatbot attributes. While a highly useful chatbot may provide a good user experience, this experience can be strengthened even further by mindful inclusion of content and chatbot characteristics that are perceived as pleasant, evocative or playful.

3. *User reports are valuable* Understanding users' experiences of chatbots is challenging. Nevertheless, insight into such experiences is critical for chatbot service providers to strengthen chatbot uptake in the general population. Through the presented research approach, we have demonstrated the feasibility and the benefit of gathering user reports through a questionnaire study based on the critical incident technique. We hope this serves to exemplify how service providers can approach gathering qualitative user reports that provide much-needed rich insight into chatbot user experience.

4. *Different users have different needs* The natural language interaction entailed in chatbots makes them highly suited for personalization. This is exemplified in our finding that pragmatic and hedonic chatbot attributes seem to have different importance for chatbot users of different age groups. The opportunity for personalization of chatbots, however, has not been sufficiently realized. Rather, current chatbots typically exhibit the same personality and provide the same content regardless of the characteristics of the user. Chatbot service providers may benefit from investing in understanding the needs of their different user groups and setting up chatbots that can adapt accordingly.

## Limitations and future work

This study contributes insight into drivers of good and poor chatbot user experience by gathering data from users of a highly diverse set of chatbots. As such, the presented findings will be of interest for chatbot research and development.

At the same time, the study has important limitations. First, whereas the study allowed us to identify key aspects of chatbot user experience, the study sample was not sufficient for a detailed breakdown of the relative importance of these aspects for smaller user groups or for different chatbots or chatbot types. This would make an interesting challenge for

future survey studies with larger samples. Specifically, we foresee more in-depth investigations into the relative differences between task-oriented chatbots and chatbots oriented towards social interaction with respect to how pragmatic and hedonic attributes can be leveraged to improve user experience.

Second, whereas the study involved a broad range of users, they were all from the same geographical region and language area. Replicating the study in other regions may lead to different outcomes—among other things because there is substantial variation in chatbot availability and support across languages. It may also be argued that the participants in this study are relatively more innovative and open than most other users with regard to their technology use, given that chatbots are still an emerging technology. Specifically, what Rogers [40] refers to as early adopters may be more prominent in this sample than they would be in a population where there is widespread uptake of chatbots. As such, generalizing from the study findings to future populations of chatbot users requires consideration of their similarities and differences with the population under study. We anticipate future research on chatbot user experience from the perspective of the pragmatic-hedonic framework, which may shed light on how user experience of chatbots evolve over time, regions and populations.

Third, whereas user experience is a subjective phenomenon, it also impacts user behaviour. Hence, a comprehensive study of chatbot user experience would benefit from also including chatbot user behaviour. We anticipate studies that combine large-scale questionnaire studies with data collection on user behaviour, for example by using log data.

# Conclusion

In this paper, we have presented a study investigating chatbot user experience. The study was conducted as a questionnaire survey, where positive and negative experiences with chatbots were gathered as free-text descriptions and were processed in a content analysis. The analysis allowed us to detail how pragmatic and hedonic attributes of user experience can lead to positive or negative user experiences. This is a relevant contribution to chatbot research and practice, as it suggests the benefit of strategically combining pragmatic and hedonic attributes while at the same time accentuating the potential risks associated with the two groups of attributes.

The study findings have also been discussed with regard to differences between task-oriented chatbots and chatbots oriented towards social interaction. Whereas the former type will need to take a starting point in pragmatic user experience attributes and then enhance the user experience with hedonic attributes, the latter may benefit from enhancing an engaging and immersive user experience by also leveraging pragmatic attributes.

As an early study of user experiences of chatbots with a theoretical basis in a pragmatic-hedonic framework of user experience, this study has limitations, as discussed above. Nevertheless, we hope that the presented findings will serve as useful steppingstones for future research and development in the direction of useful and pleasurable chatbot user experiences.

## Compliance with ethical standards

**Conflict of interest**  On behalf of all authors, the corresponding author states that there are no conflicts of interest.

## References

1. Avula S, Chadwick G, Arguello J, Capra R (2018) Searchbots: user engagement with chatbots during collaborative search. In: Proceedings of the 2018 conference on human information interaction & retrieval. ACM, New York, NY

2. Bernhaupt R (ed) (2010) Evaluating user experience in games: concepts and methods. Springer, Berlin

3. Bhandari U, Neben T, Chang K, Chua WY (2017) Effects of interface design factors on affective responses and quality evaluations in mobile applications. Comput Hum Behav 72:525–534

4. Brandtzaeg PB, Følstad A (2018) Chatbots: Changing user needs and motivations. Interactions 25(5):38–43

5. Brandtzaeg PB, Følstad A (2017) Why people use chatbots. In: Proceedings of the international conference on internet science. Springer, Cham, Switzerland, pp 377–392

6. Chaves AP, Gerosa MA (2018) Single or multiple conversational agents?: An interactional coherence comparison. In: Proceedings of the 2018 CHI conference on human factors in computing systems. ACM, New York, NY, paper no. 191

7. Chen H, Liu X, Yin D, Tang J (2017) A survey on dialogue systems: recent advances and new frontiers. ACM SIGKDD Explor 19(2):25–35

8. Crutzen R, Peters GJY, Portugal SD, Fisser EM, Grolleman JJ (2011) An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 48(5):514–519

9. Dale R (2016) The return of the chatbots. Nat Lang Eng 22(5):811–817

10. Dell N, Vaidyanathan V, Medhi I, Cutrell E, Thies W (2012) "Yours is better!": participant response bias in HCI. In: Proceedings of the SIGCHI conference on human factors in computing system. ACM, New York, NY, pp 1321–1330

11. Deriu J, Rodrigo A, Otegi A, Echegoyen G, Rosset S, Agirre E, Cieliebak M (2019) Survey on evaluation methods for dialogue systems. arXiv preprint. arXiv:1905.04071

12. Diefenbach S, Hassenzahl M (2011) The dilemma of the hedonic—appreciated, but hard to justify. Interact Comput 23(5):461–472

13. Drift (2018) The 2018 state of chatbots report. Technical report, Drift. https://www.drift.com/blog/chatbots-report/

14. Ezzy D (2002) Qualitative analysis: practice and innovation. Routledge, London

15. Field A (2013) Discovering statistics using IBM SPSS Statistics, 4th edn. SAGE, London, UK

16. Fitzpatrick KK, Darcy A, Vierhile M (2017) Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. JMIR Ment Health 4(2):e19

17. Flanagan JC (1954) The critical incident technique. Psychol Bull 51(July):327–358

18. Forlizzi J, Battarbee K (2004) Understanding experience in interactive systems. In: Proceedings of the 5th conference on designing interactive systems. ACM, New York, NY, pp 261–268

19. Fryer LK, Ainley M, Thompson A, Gibson A, Sherlock Z (2017) Stimulating and sustaining interest in a language course: an experimental comparison of chatbot and human task partners. Comput Hum Behav 75:461–468

20. Følstad A, Brandtzaeg PB (2017) Chatbots and the new world of HCI. Interactions 24(4):38–42

21. Gremler DD (2004) The critical incident technique in service research. J Serv Res 7(1):65–89

22. Hall E (2018) Conversational design. A Book Apart, New York, NY

23. Hassenzahl M (2018) The thing and I: understanding the relationship between user and product. In: Blythe M, Monk A (eds) Funology 2. Human–computer interaction series. Springer, Cham, pp 301–313

24. Hassenzahl M, Tractinsky N (2006) User experience—a research agenda. Behav Inf Technol 25(2):91–97

25. Hassenzahl M (2008) User experience (UX): towards an experiential perspective on product quality. In: Proceedings of the 20th Conference on l'Interaction Homme-Machine. ACM, New York, NY, pp 11–15

26. Hassenzahl M, Burmester M, Koller F (2003) AttrakDiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In: Szwillus G, Ziegler J (eds) Mensch & Computer 2003: Interaktion in Bewegung. Vieweg+ Teubner Verlag, pp 187–196

27. Hill J, Ford WR, Farreras IG (2015) Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations. Comput Hum Behav 49:245–250

28. Hornbæk K, Hertzum M (2017) Technology acceptance and user experience: A review of the experiential component in HCI. ACM Trans Comput-Hum Interaction (TOCHI) 24(5):33@@

29. ISO (2010) ISO 9241-210:2010 (E). Ergonomics of human system interaction—part 210: human-centred design for interactive systems. International Organization for Standardization (ISO), Geneva, Switzerland

30. Jurafsky D, Martin JH (2019) Speech and language processing. Draft of October 2, 2019. https://web.stanford.edu/~jurafsky/slp3/26.pdf

31. Khatri C, Hedayatnia B, Venkatesh A, Nunn J, Pan Y, Liu Q, Song H, Gottardi A, Kwatra S, Pancholi S, Cheng M, Chen Q, Stubel L, Gopalakrishnan K, Bland K, Gabriel R, Mandal A, Hakkani-Tur D, Hwang G, Michel N, King E, Prasad R (2018) Advancing the state of the art in open domain dialog systems through the Alexa Prize. arXiv preprint. arXiv:1812.10757

32. Kinsella B, Mutchler A (2019) Smart speaker consumer report, March 2019. Technical report, Voicebot & Voicify. https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/

33. Kujala S, Roto V, Väänänen-Vainio-Mattila K, Karapanos E, Sinnelä A (2011) UX curve: a method for evaluating long-term user experience. Interact Comput 23(5):473–483

34. Law EL-C, Van Schaik P, Roto V (2014) Attitudes towards user experience (UX) measurement. Int J Hum-Comput Stud 72(6):526–541

35. Law EL-C, Roto V, Hassenzahl M, Vermeeren AP, Kort J (2009) Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the SIGCHI conference on human factors in computing system. ACM, New York, NY, pp 719–728

36. Lowe R, Noseworthy M, Serban IV, Angelard-Gontier N, Bengio Y, Pineau J (2017) Towards an automatic Turing test: learning to evaluate dialogue responses. In: Proceedings of the 55th annual meeting of the Association for Computational Linguistics. ACL, Stroudsburg, PA, pp 1116–1126. https://www.aclweb.org/anthology/P17-1103/

37. Luger E, Sellen A (2016) Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In: Proceedings of the SIGCHI conference on human factors in computing system. ACM, New York, NY, pp 5286–5297

38. Nordheim CB, Følstad A, Bjørkli CA (2019) An initial model of trust in chatbots for customer service—findings from a questionnaire study. Interact Comput 31(3):317–335

39. Porcheron M, Fischer JE, Reeves S, Sharples S (2018) Voice interfaces in everyday life. In: Proceedings of the 2018 CHI conference on human factors in computing systems. ACM, New York, NY, paper no. 640

40. Rogers EM (2010) Diffusion of innovations, 5th edn. Free Press, New York

41. Shevat A (2017) Designing bots: creating conversational experiences. O'Reilly Media, Sebastapol

42. Shum HY, He XD, Li D (2018) From Eliza to XiaoIce: Challenges and opportunities with social chatbots. Front Inf Technol Electron Eng 19(1):10–26

43. Thies IM, Menon N, Magapu S, Subramony M, O'Neill J (2017) How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. In: IFIP conference on human–computer interaction. Springer, Cham, Switzerland, pp 441–459

44. Tractinsky N, Katz AS, Ikar D (2000) What is beautiful is usable. Interact Comput 13(2):127–145

45. Van Schaik P, Ling J (2008) Modelling user experience with web sites: usability, hedonic value, beauty and goodness. Interact Comput 20(3):419–432

46. Walker MA, Litman DJ, Kamm CA, Abella A (1997) PARADISE: a framework for evaluating spoken dialogue agents. In: Proceedings of 35th annual meeting of the Association for Computational Linguistics and 8th conference of the European chapter of the Association for Computational Linguistics. ACL, Stroudsburg, PA, pp 271–280. https://www.aclweb.org/anthology/P97-1035/

47. Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 9(1):36–45