Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots

Nina Svenningsson*
Kristianstad University
Kristianstad, Sweden
nina.svenningsson@gmail.com

Montathar Faraon
Kristianstad University
Kristianstad, Sweden
montathar.faraon@hkr.se

ABSTRACT

Artificial intelligence (AI) is gaining traction in service-oriented businesses in the form of chatbots. A chatbot is a popular type of social AI that uses natural language processing to communicate with users. Past studies have shown discrepancies in terms of whether or not a chatbot should communicate and behave like a human. This article aims to explore these discrepancies in order to provide a theoretical contribution of a list of factors related to perceived humanness in chatbots and how these may consequently lead to a positive user experience. The results suggest that a chatbot should have the following characteristics: avoiding small talk and maintaining a formal tone; identifying itself as a bot and asking how it can help; providing specific information and articulating itself with sophisticated choices of words and well-constructed sentences; asking follow-up questions during decision-making processes and; providing an apology when the context is not comprehensible, followed by a question or a statement to dynamically move a conversation forward. These results may have implications for designers working in the field of AI as well as for the wider debates and the broader discourse around the adoption of AI in society.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design** → **Interaction design theory, concepts and paradigms**
• **Computing methodologies** → **Artificial intelligence** → **Intelligent agents; Discourse, dialogue and pragmatics**

KEYWORDS

artificial intelligence; AI; chatbot; humanness; conversation; user experience; design guideline

1 INTRODUCTION

Artificial intelligence (AI) is the intelligence that is demonstrated by machines [17]. The first definition of AI is attributed to John McCarthy, who is often recognized as the father of AI; he coined the term and defined it as "the science and engineering of making intelligent machines" [29]. Later definitions of AI include "the intelligence demonstrated by machines, in contrast to the natural intelligence (NI) displayed by humans and other animals. AI [...] implements human cognitive functions such as perceiving, processing input from the environment and learning" [17, p. 73]. Similar definitions of AI have been put forward by other researchers, for instance "machines that exhibit aspects of human intelligence" [20, p. 155] and a rational agent "that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome" [46, p. 4]. For many people, the thought of AI is a dystopian one, leading to anxiety and concerns about what will happen to mankind if AI becomes too intelligent [28]. However, AI is already being used by most of us in a range of areas every day [19, 32], for example AI that filter spam from email inboxes (e.g., Gmail), video streaming services that recommend what to watch next (e.g., Netflix), and web applications that identify people in photographs [19].

The emergence of social media, such as Facebook and Twitter at the beginning of the 21st century, has been a game-changing phenomenon in the field of online communication. Within this field, a growing interest in the development of digital chatbots (henceforth chatbots) has been observed [9, 14]. A chatbot could be described as a social AI that is designed to perform conversations with humans [56] by communicating through natural language in order to achieve a certain result [9], for example, aiding users with customer support or assisting with booking flights. According to some predictions, 50% of companies will spend more money on creating chatbots than traditional apps in 2021 (i.e., apps that are downloaded from an online store to a mobile device) [41].

By and large, users today do not feel that their expectations of what a chatbot should be capable of are being met; for instance, a chatbot may not understand their questions or intentions [21], and human-like attributes create expectations that are higher than the chatbots can meet [27]. Users also have difficulty grasping what a chatbot is capable of, because they do not receive enough feedback to build a correct image of its capabilities [21, 27].

In the process of designing chatbots, some argue that a chatbot should be given human-like behavior [2, 40]; however, there is a lack of consensus on whether this is desirable [37, 43, 48]. On one hand, giving a chatbot a personality is widely viewed as an important part of its design, as it appeals to users in terms of creating engagement, encouraging playfulness, and building trust [21, 25, 33]. On the other, humorous chatbots can falsely lead users to believe that they are smarter and more capable than they are [27, 30], and a chatbot personality that is too human-like can be perceived by users as disturbing and uncanny [27, 30, 33–35, 49].

Based on the aforementioned reasoning, this article aims to explore and identify the factors that can contribute to creating a human-like user experience (UX) with a chatbot, and to formulate

*Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
AICCC 2019, December 21–23, 2019, Kobe, Japan
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7263-3/19/12.
DOI: https://doi.org/10.1145/3375959.3375973

RIGHTSLINK

a design guideline that may aid the future design of chatbots. In this context, UX can be defined as "the experience of using an interactive system" [24]. Correspondingly, this article is guided by the following research question: *What factors contribute to the creation of a human-like UX of a chatbot, and how can a design guideline be formulated to support the future design of chatbots?* To address this question, a quantitative survey was used, which will be further described in the methodology section. The rest of this article is organized as follows. The next section presents a theoretical review of AI, chatbots, and factors related to perceived humanness in chatbots. Then, our methodology is described with a focus on the quantitative approach. Using this approach, the reader is then presented with the results and a corresponding analysis. Based on the findings, a design guideline is proposed for future consideration when designing a chatbot. The final section concludes by discussing the results in comparison with related research and providing suggestions for future work.

2 LITERATURE REVIEW

2.1 Artificial intelligence

The term AI was coined in the 1950s, and during this decade Alan Turing published an article which for many scientists marks the birth of AI [17, 51]. The article proposed the Turing test, which aims to measure the intelligence of machines and to determine whether they could convince a jury that they are in fact human. There are doubts regarding to which extent this could indicate that a machine is in fact intelligent, and even Turing emphasized that the focus of the test was for the machines to *imitate* a human being [5]. It was not until 2014 that a machine managed to pass the Turing test [55]. The chatbot, referred to as Eugene Goostman, managed to convince more than one third of the judges that it was a human. It is not easy for AI to imitate human intelligence, as there are many aspects to take into consideration (e.g., cognitive, social and psychological factors) [53, 55]. These factors are dependent on the context, and differ between situations [53].

AI technologies have the potential to complement and support tasks that are performed by humans today [28]. On one hand, if AI is granted human-equivalent rights and extensive access to decision-making processes, this could lead to AI taking control and consequently result in human beings becoming secondary [28], and this has raised concerns in terms of ethics, transparency and accountability [15, 50]. On the other hand, AI has the potential to revolutionize society and to create a utopia in which human beings could cure diseases [16], AI could supplement human contributions [22], and individuals could spend more time pursuing personal growth and individual passions. Today, AI has been used to automate many parts of our lives, for instance in autonomous vehicles (e.g., Tesla), robots for home cleaning (e.g., Xiaomi Roborock), digital nurses that monitor the condition of a patient (e.g., Sensely), and voice-controlled speakers (e.g., Apple Homepod) to manage home accessories (e.g., to switch on lights or play music). The trends in AI suggest that human beings could be replaced not only in professions that require mechanical intelligence (e.g., call center agents, waiters, taxi drivers) but also professions that require analytical and intuitive intelligence (e.g., accountants, analysts, managers) [17, 20]. Even so, there are doubts that AI could ever become a serious concern (i.e., replacing humans); it has been suggested that human consciousness may not be able to be summed up in rules, and machines can only be programmed to follow rules [28]. This could be interpreted as meaning that it may be difficult to design and create an AI that fully mirrors human-like intelligence.

A chatbot is one example of AI that uses Natural Language Processing (NLP) to communicate and interact with humans [56]. NLP is a combination of techniques that allows for the analysis of text or language on one or several levels [47]. Through NLP, human beings can interact with AI using natural language. In some cases, sophisticated models such as the adaptive response model are being used to interpret what human beings are saying and to provide adaptive responses. This model has been shown to increase perceived humanness and faster responses [47].

Chatbots use the same kind of dialogue patterns as in SMS and messaging applications (e.g., Facebook Messenger), which are growing in popularity [9, 14]. Since 2014, companies and developers have created chatbots inside these applications, which has increased the accessibility of chatbots [23, 58].

2.2 Chatbots

A chatbot could be defined as a social AI that has been designed to carry out conversations with humans using NLP [9, 56]. The flexibility this creates has led to chatbots being used in various fields [2, 48], for instance in customer service [6] (e.g., ChatBot, chatbot.com), libraries [32] (e.g., Pixel, pixel.unl.edu), and health care [42] (e.g., MedWhat, medwhat.com). Chatbots can be used for various purposes, such as helping users to find information, supporting collaboration between users, and automating parts of a workflow that are repetitive [40].

Some researchers have suggested a classification of interaction styles of chatbots that divides chatbots into four categories: (i) *dull*, in which a chatbot answers with single words and often repeats the same or similar phrases; (ii) *alternate vocabulary*, in which a chatbot has a larger base of variations of the same answers; (iii) *relationship building*, in which vocabulary and topics can change between easygoing and professional, and a chatbot can control the flow of conversation by giving spontaneous information or making a joke; and (iv) *human-like*, where a chatbot learns from previous experience and uses past knowledge to interact using more subtle conversation patterns and meaningful dialogue [40]. Due to difficulties in mapping the capabilities of chatbots, users are often not able to tell which category a chatbot belongs to, and by extension what they can expect from it [21, 27]. This leads to high expectations, and disappointment and frustration when these expectations cannot be met. It is not easy to predict how users will react during interactions with chatbots, as this is highly dependent on the user's personality [26, 33].

One study have identified the following factors that contribute to successful interactions with AI: *clear and uncomplicated answers*, *relevant answers* and *answers that are not too short, but also not too long* [6]. In this study, artificial conversations were evaluated using Gricean maxims (e.g., quality, quantity, relation), which describe the principles that enable effective communication. For instance, the quantity maxim states that the "speaker's utterance provides as much information as appropriate, not more, not less," while the

RIGHTSLINK

relation maxim states that the "speaker's utterance is relevant to the context and the topic of the conversation" [6].

Similarly, in a dialogue with a chatbot, previous research has indicated that users appreciate the use of *sophisticated words* (i.e., the use of a diverse vocabulary) and *well-constructed sentences* (i.e., complex language) [26]. These factors, in combination with lengthy dialogue, were connected to immediacy and were perceived as an indicator of care [54]. In addition, high grammatical quality was associated with positive evaluations of humanness, and was related to the perception of being human-like [26].

Previous research also confirms that users expect chatbots to behave in a human-like manner [58]. When first-time users were asked about the expectations they had of a chatbot, the following four categories were indicated as important: *high performing*, *smart*, *seamless*, and *personable*. The first, *high performing*, refers to being fast, efficient, and reliable. In general, a chatbot is meant to show a user the fastest and most efficient way to complete a task. The second, *smart*, is a measure of how knowledgeable, accurate, and predictive a chatbot is. For example, a chatbot could suggest a time of a departure that depends on the location of an appointment. The third, *seamless*, focuses on characteristics such as the smoothness, ease, and flexibility of the experience (i.e., "all in one") and the ability to adjust to different circumstances, for instance a task that requires different actions on different days. The fourth, *personable*, refers to the understanding of the user and to the aspect of how likeable a chatbot is [58].

In line with these aspects, users have indicated that they expect a chatbot to engage in a dynamic conversation, for example in which it asks relevant follow-up questions in order to steer the conversation in the right direction and to attempt to find the correct answer [21, 47]. Users also expected a chatbot to engage in small talk and to have a personality that matches the context for which it was designed. Several studies have shown that users have high expectations of the capabilities of chatbots, and these expectations are currently not being fulfilled [4, 21, 27].

2.3 Designing a chatbot with humanness

One example of a framework that is used today when developing autonomous agents with human-like behavior is the BDI model [38, 39]. The BDI model is based on three concepts: *beliefs*, *desire*, and *intention* [38, 45]. Agents make decisions based on "beliefs which represent the individual's knowledge about the environment and about their own internal state; desires or more specifically goals (non-conflicting desires which the individual has decided they want to achieve); and intentions which are the set of plans or sequence of actions which the individual intends to follow in order to achieve their goals." [1, p. 212]. The BDI model has proved to be successful in creating human-like characters [39]. At the same time, it is an abstraction of human deliberation and the model lacks many generic aspects of human behavior and reasoning [39, 45].

When designing a human-like chatbot, it may be essential to consider which factors can influence the UX to fulfill the user's needs. UX can be defined as "the experience of using an interactive system" [24]. To elaborate on this meaning, UX may be described by the following three components: (i) its *usefulness* - if, or how well, an artifact enables a user to reach a certain goal; (ii) its *usability* - how easy to use an artifact is perceived to be; and (iii) *satisfaction* - how satisfied a user is with the interaction with an artifact [13].

One aspect of UX in the context of AI is how human-like chatbots should behave, and how this affects UX. Although previous studies have explored how chatbots could evolve to converse or behave in a human-like manner [2, 21, 27, 30, 33, 56], there is no consensus on whether this is the correct way to move forward from a UX perspective of interacting with chatbots [37, 43, 48].

The simulation of feelings in chatbots can be engaging, but people are also sensitive towards behaviour that seems "off". This means that if an emotional response from a chatbot does not match the conventions and unwritten rules of human conversation, this may create discomfort in users [30]. There is also a risk that signs of emotional intelligence in chatbots can lead to users interpreting them as being more capable than they actually are, leading to high expectations that cannot be met [27, 30]. The users could also interpret the chatbots as being more conscious in producing their answers [57]. Emotional responses that are relevant to a situation have been shown to create a trustworthy impression, and users perceive chatbots that are capable of performing such responses as more competent and warmer [10].

It has been argued that a chatbot can become more trustworthy if it is more robotic than human-like [56]. A machine does not have to be human-like to be engaging [33], and there is not necessarily a correlation between how human-like and emotionally intelligent a chatbot is, and how likeable it is [30]. In some cases, human-like chatbots have even created frustration when engaging with users [43]. However, studies have shown that users are attracted to human-like chatbots [2] and AI that allows for a human-like use of language [21]. Whether a chatbot should imitate human behavior or have their own personality could be dependent on context [37].

Closely related to previous considerations is the aspect of creating trust in chatbots, which in part depends both on the design of personality and on the design of the conversation. Factors that can contribute to users trusting chatbots are: (i) *small talk* [3, 44]; (ii) *a chatbot being open and transparent about being a machine* [33]; (iii) *the ability to react and respond to signals that indicate emotional and social behavior* [30]; and (iv) *an authentic chatbot that matches the user on a personal level and in conversation* [37]. In this case, authentic means that the chatbot is aware of and capable of remaining in context, that it is transparent and can act as a person with values, attitude and culture [37].

Finally, it has been recommended taking the following factors into consideration when designing a chatbot: (i) *clarify the chatbot's competence and abilities*; (ii) *have the ability to maintain the context throughout a conversation*; (iii) *enable small talk*; (iv) *manage failed dialogue*; and (v) *end a conversation in an appropriate way* [21]. It has been shown that users preferred chatbots that conversed in a human-like manner (e.g., remaining within context, understanding negative statements, managing failure and asking intelligent questions) and that took advantage of familiar turn-taking conversation roles, as seen in messaging conversations [21].

2.4 Summary

Previous studies have identified factors that have been shown to influence how human-like a chatbot is perceived to be. These are

RIGHTSLINK

as follows: (i) *small talk* [3, 44]; (ii) *way of conduct in tasks* [21, 48]; (iii) *adhering to context* [21, 37]; (iv) *natural flow of conversation* (e.g., turn-taking, pauses) [21, 48]; (v) *sophisticated choices of words and well-constructed sentences* [26]; (vi) *a dynamic approach* (e.g., asking follow-up questions) [21, 47]; (vii) *expressing feelings based on context* [10, 21]; and (viii) *reacting to social and emotional cues* [30]. In the next section, the methodology of this study will be described, with a focus on the survey that was conducted.

3 METHODOLOGY

3.1 Sampling

An assessment was made with regard to how feasible it was to include each of the previously mentioned factors in the survey. For example, it was challenging to include the third factor, namely adhering to context, since this would have required a working chatbot. Based on this consideration, factors three, four, and eight were excluded, resulting in the following five factors: (F1) *small talk*; (F2) *way of conduct in tasks*; (F3) *sophisticated choices of words and well-constructed sentences*; (F4) *dynamic approach* (e.g., asking follow-up questions); and (F5) *expressing feelings based on context* (e.g., frustration). A quantitative survey was applied, including questions where participants were presented with several situations. These situations were based on factors F1–F5 and were related to user experience as well as human-like behavior that had been identified from the literature review. All questions were self-generated and created specifically for this survey.

3.2 Participants

A total of 75 participants completed the survey. After data processing to exclude outliers and incomplete values, 69 participants remained. A total of 47% were females (n = 32), 52% were males (n = 36), and 1% (n = 1) did not wish to disclose their gender. Participants had an age range of 20 to 61 years (M = 35.89; SD = 9.56). With regard to how many hours they spent on the Internet per day, 51% indicated 2–5 hours (n = 35), 38% answered 6–10 hours (n = 26), 10% responded 11–20 hours (n = 7), and 1% was online for 0–1 hours (n = 1). 99% indicated that they used a smartphone (n = 68) and 1% did not (n = 1).

3.3 Materials

The survey consisted of two sections (see Appendix A and B). The first section aimed to collect demographic data about the participants, such as age, gender, how much time they spent on the Internet and whether they used smartphones. It also focused on gathering information related to how often and in what situations participants had encountered chatbots (e.g., searching for information, customer service). A definition of a chatbot was included in this section with the purpose of supporting participants in their understanding of what a chatbot is. In addition, the first section also examined whether participants looked for chatbots when they needed help on websites. This section was included with the purpose of better understanding the extent to which a chatbot is a desirable and useful tool. The second section consisted of 11 questions, specifically generated for this survey, in which participants were requested to evaluate how they wished to be addressed by a chatbot and how it should articulate itself during the conversation. In the following, each of the 11 questions will be briefly described in relation to the mentioned factors (F1–F5).

The first question aimed to investigate how users wished to be addressed by a chatbot on a personal level, i.e., who would initiate the conversation and whether a user wished to be greeted by name (F1). The second question examined how likely a user would be to continue interacting with a chatbot if it did not understand what the user wanted, and the ways in which the chatbot should describe how to move forward from such a situation (F5). The third question explored the degree to which it is desirable that a chatbot has a personality and the extent to which a chatbot should share it through small talk (F1). The fourth question focused on how detailed the information in a chatbot's responses should be (F3). The fifth question considered whether a user would ask for follow-up questions to determine whether a chatbot had understood a given question and could give a relevant response to it (F4). The sixth question sought to understand what behavior is desirable in a formal context (e.g., a bank) and whether a user expected similar behavior to that of a physical visit (F2). The seventh question aimed to study which behavior is desirable in an informal context (e.g., entertainment/playing music) and whether a user expected a chatbot to have an entertaining personality (F2). The eighth question focused on how specifically a user wishes a chatbot would articulate itself with regard to what it is capable of and the context in which it is operating (F2). The ninth question examined the parlance a chatbot should use in terms of well-constructed sentences (F3). The tenth question probed how desirable it was that a chatbot could interpret situations in which a user is frustrated, and the degree to which it should show apologetic behavior (F5). The eleventh question looked more closely at how desirable it is that a chatbot asks follow-up questions to ensure a user is provided with relevant responses, and ultimately to lead to a successful resolution of a conversation (F4).

A reliability test was conducted to examine the internal consistency of these questions. The test showed a Cronbach's alpha value of 0.59, which indicated satisfactory internal consistency [8].

3.4 Procedure

The survey was created and conducted with the Emarketeer tool. It was sent via email to employees in a company that worked with chatbots and their customers. In addition, the survey was also shared via social media (LinkedIn and Facebook) to reach a wider audience who may have had experience with chatbots. The survey was active over a period of 10 days. Upon opening the link to the survey, participants were briefed with the purpose of the survey and informed that it would take approximately 5–10 minutes to complete. Once the survey was closed, data were exported to Excel and imported into SPSS for further processing. Questions related to a particular factor (F1–F5) were analyzed together.

3.5 Ethical considerations

Prior to data collection, ethical considerations were taken into account in accordance with the guidelines set forth by the Swedish Research Council [7]. Upon opening the survey, information was given to participants that their participation was on a voluntary and confidential basis, and all responses were anonymized. They were

RIGHTSLINK

informed that they could end their participation at any time during the survey. Furthermore, participants were given information that the collected data would only be used for research purposes within the scope of this study and not shared with any third parties.

4 RESULTS AND ANALYSIS

4.1 Demographics

Participants indicated the following regarding their experience of interacting with chatbots: 7% had not interacted with a chatbot (n = 5), 49% a few times a year (n = 34), 20% a few times a month (n = 14), 15% a few times a week (n = 10), and 9% every day (n = 6). The results showed that participants had interacted with a chatbot, with the majority having an interaction a few times per year. Concerning when participants encountered chatbots, they indicated that this was in the following situations: searching for information (n = 29), checking for the weather forecasts (n = 13), playing music (n = 17), customer service (n = 43), survey/evaluation (n = 15). The remaining options (i.e., booking meetings, to have someone to talk to, to get inspiration when shopping online, job-related situations, none of the above) were selected by five participants or fewer.

Since chatbots are not widely utilized, it was expected that the results would show that the majority of participants encountered chatbots in a customer service situation and when searching for information. Moreover, participants were requested to indicate whether they would ask for help via chat on websites: 23% said never (n = 16), 68% indicated sometimes (n = 47), and 9% answered all the time (n = 6). The results revealed that the majority of participants were inclined to ask for help using chat on websites, which may be an indicator that chatbots could offer an increasingly essential way of providing customers with instant support. In the following section, the results and analyses of the survey are presented.

4.2 Survey

4.2.1 Small talk. This factor (F1) was investigated in question one and three. The first question was designed to examine the degree to which a chatbot should be formal when greeting a user, and the third was concerned with the type of information a chatbot should share in relation to a task, which in this case was looking for a destination on a business trip.

A chi-square test was conducted for the first question to determine whether there were statistically significant differences in preferences of how participants wanted to be greeted by a chatbot. The test was statistically significant, $\chi^2(3, N = 69) = 19.99$, $p < .001$. Further investigation revealed a division between the majority of participants, in which the largest group (39.1%) stated that it should be their initiative to start a conversation, and the second largest group (37.7%) indicated that they wished to be greeted by name and asked what they want help with. Two smaller groups indicated that they only wished to be greeted by saying hello (13%) or to be greeted by name (10.1%). These results suggested that the second largest group may have valued the aspect of being asked what they needed help with.

Likewise, a chi-square test was conducted for the third question, which concerned the information that participants preferred to receive via chatbot when looking for a business trip destination. Two of the six options were excluded because they had less than five participants, which was one of the assumptions of the test that needed to be fulfilled. The test was statistically significant, $\chi^2(3, N = 66) = 70.36$, $p < .001$. The findings of the third question showed a strong preference among participants to receive a formal update while a chatbot was processing a request (66.7%). By and large, the results indicated that small talk may not be preferred when interacting with a chatbot. However, the results could be context-dependent in the sense that small talk may not be preferred in a service-oriented context but could be accepted in a social context.

4.2.2 Way of conduct in tasks. This factor (F2) was examined in questions six, seven, and eight. All three questions investigated the opening message that participants preferred for different tasks, which included banking, playing music and booking a meeting. The sixth question focused on banking, and prior to a chi-square test, one of four options was excluded because it did not meet the minimum requirement of five participants. The test was statistically significant, $\chi^2(2, N = 67) = 20.54$, $p < .001$. The largest group (47.8%) preferred that a chatbot should identify itself as a bot, followed by asking how it could help. The second largest (42%) preferred to be asked how a chatbot could help. A smaller group (7.2%) wished to be asked if they wanted to spend or save. By and large, the results seem to indicate that participants wanted to be asked how they could be helped, with some preferring a chatbot to identify itself as a bot. Furthermore, the results indicated that participants did not prefer a chatbot to express itself in a humorous or informal manner.

The seventh question related to a chatbot assisting when playing music. A chi-square test for the question was statistically significant, $\chi^2(3, N = 69) = 39.46$, $p < .001$. The largest group (53.6%) indicated that they wished to be asked what they wanted to listen to, and the second largest (30.4%) responded that they wanted to be asked for a genre to listen to. In addition, the results suggested that the majority of participants preferred neither a formal response (e.g., "please state a genre") (7.2%) nor an informal one (e.g., "groovy, you're back!") (8.7%), but rather a balanced response.

Question eight considered the interaction with a chatbot when booking a meeting. One of the five options had less than five participants, and was therefore excluded. A chi-square test was conducted and showed no statistical significance. However, further investigation showed an equal division between the two largest groups. The first (30.4%) preferred to be asked whether to book a room or manage their calendar, while the second (30.4%) wanted a chatbot to identify itself as a bot and to show available meeting times based on their calendar. Overall, the results suggest that it may not be desirable for a chatbot to have an entertaining personality.

4.2.3 Sophisticated choice of words and well-constructed sentences. Factor three (F3) was investigated in questions four and nine. The questions were formulated based on correct punctuation, word choice, and sentence length. The common denominator for both questions was the context of travel. The fourth question focused on providing timetable information for bus departures, and the ninth information about a booked trip. A chi-square test for the fourth question was statistically significant, $\chi^2(3, N = 69) = 29.73$, $p < .001$. A closer look at the results showed that the largest group (43.5%) preferred specific information that included bus numbers, destinations, and timetables. The second largest group (39.1%) was satisfied with only receiving specific information about bus

RIGHTSLINK

numbers and timetables. In general, and in line with the results of the seventh question, the majority of the participants preferred neither a short non-specific response (e.g., "12:15") (7.2%) nor a long one (e.g., "there are five different buses going to the train station from your bus stop...") (10.1%), but rather a balanced response.

The ninth question addressed the responses that participants wished to receive when booking a trip. A chi-square test was performed for the ninth question and was statistically significant, $\chi^2(2, N = 69) = 18.35$, $p < .001$. In general, the largest group (55.1%) preferred the option with the longest sentence, the most sophisticated choice of words, and a well-constructed sentence structure over the other shorter, less sophisticated, less structured options. Nonetheless, the second largest group (31.9%) selected an option with less sophistication and structure. The smallest group (13%) chose the shortest and the least sophisticated, well-structured option. One explanation behind the choice of the second group may be that this option was shorter but contained the same vital information as the option selected by the largest group. Another explanation is that in the era of social media, participants may expect to select and process shorter messages.

4.2.4 Dynamic approach. The fourth factor (F4) was analyzed by means of questions five and eleven. These questions inquired into whether participants preferred to receive follow-up questions in specific contexts, such as asking about the weather and making a purchase choice of a hair trimmer. The fifth question asked participants to indicate their preferred response when getting weather updates. Prior to a chi-square test, two of the four options were excluded as they had fewer than five participants. The test was statistically significant, $\chi^2(1, N = 67) = 32.97$, $p < .001$. The results showed that participants did not prefer a follow-up question (2.9%); instead the majority (82.6%) preferred to get the following detailed information about the weather: (i) *their current city based on geolocation*; (ii) *temperature in degrees in Celsius*; and (iii) *whether there would be any clouds*. A small group (14.5%) preferred to get the same detailed information but without the current city.

The eleventh question looked into follow-up questions when making a purchase choice. Two of the four options for the questions were excluded as they had fewer than five participants. Following this, a chi-square test was performed and was not statistically significant. The results showed an equal division between one group (44.9%) who wanted to get a purchase recommendation based on reviews, and a second group (46.4%) who indicated that they wished to be asked a follow-up question to make a better purchase choice. These results may be interpreted as showing that participants wanted to give more specific information to make a better purchase choice, while also wanting to get more specific information about purchase choices in terms of reviews. Overall, the results seem to suggest that when asking a chatbot for information that does not involve making a choice, no follow-up questions may be required. Instead, follow-up questions may be preferable when making a purchase, where more information should be both asked for and provided to inform the decision-making process. It may be a good idea for a chatbot to foster common understanding, and this could be done by asking follow-up questions or repeating the vital elements of a question in the answer.

4.2.5 Expressing feelings based on context. Factor five (F5) focused on the aspect of when a chatbot does not understand a user, and was explored in questions two and ten. The second question involved responses that participants want to receive when a chatbot does not understand. Before a chi-square test was conducted for the second question, two of the five options were excluded as they had fewer than five participants for each option. The test was statistically significant, $\chi^2(2, N = 68) = 12.56$, $p < .05$. The results showed that the largest group (44.9%) preferred to receive an apology from a chatbot, followed by being asked if they wished to talk to a human instead, while the second largest group (40.6%) favored the option of receiving an apology followed by being asked to rephrase their question in a different way. A smaller group (13%) was satisfied with receiving a response in which a chatbot apologized followed by saying that it did not understand.

The tenth question looked more closely at the aspect of frustration and at how participants want a chatbot to respond when they express frustration. A chi-square test was statistically significant, $\chi^2(3, N = 69) = 16.39$, $p < .05$. Interestingly, the largest group (37.7%) pointed out that they preferred an apology and then be asked to rephrase their question, while the second largest group (34.8%) preferred an apology followed by a statement to rephrase. Two smaller groups preferred either a short response without an apology (20.3%) or a longer response with an apology (7.2%). A common factor for both questions was that participants preferred to receive an apology from a chatbot when it did not understand them, followed by being asked to rephrase their question in order to move the conversation forward. However, it was not desirable that a chatbot should apologize in a prolonged response.

4.3 Design guideline for chatbots

A common approach to reporting results within design research involves design guidelines [11, 12]. A frequently used template for formulating design guidelines is as follows [52]: "If you want to design intervention X [for the purpose/function Y in context Z], then you are best advised to give that intervention the characteristics A, B, and C [substantive emphasis], and to do that via procedures K, L, and M [procedural emphasis], because of arguments P, Q, and R". The term 'intervention' in this template refers to "products, programs, materials, procedures, scenarios, processes" [31, p. 5]. In the following, the term is rephrased in the proposed design guideline as 'a chatbot'.

The guideline is motivated and underpinned by the results of the survey, and reads as follows: *If you want to design a chatbot, then you are best advised to give that chatbot the characteristics of (a) avoiding small talk and maintaining a formal tone; (b) identifying itself as a bot and asking how it can help; (c) providing specific information and articulating itself with sophisticated choices of words and well-constructed sentences; (d) asking follow-up questions in decision-making processes; and (e) providing an apology when the context is not comprehensible followed by a question or a statement to dynamically move a conversation forward.*

5 DISCUSSION

The purpose of this study was to explore and identify factors that could contribute to creating a human-like user experience with

RIGHTSLINK

a chatbot as well as formulating a design guideline to aid in the future design of chatbots. The research question guiding this article was: *What factors contribute to the creation of a human-like UX of a chatbot, and how can a design guideline be formulated to support the future design of chatbots?* This question was addressed using a quantitative survey that resulted in five factors: *small talk*, *way of conduct in tasks, sophisticated choices of words and well-constructed sentences*, *a dynamic approach*, and *expressing feelings based on context*. Based on these factors, a design guideline was formulated.

Previous research has suggested that small talk could lead to increased trust in a chatbot [21, 44]. Other studies have shown individual differences, in which extroverted individuals reported increased trust and introverted ones did not experience any difference in trust when interacting with a chatbot [3]. The results of this study are in contrast to previous findings, and indicate that small talk may not be preferred. One reason behind this could be that the results depend on not only users' personality but also on context. Small talk may not be desired in a service-oriented context but could be appropriate in a social context.

The results also indicated that a chatbot should interact in a clear, concise way, and not involve humor. A chatbot should identify itself as a bot and should ask how it can help with different tasks. The results provide support for previous research that has identified risks in using humor [27], but contrast with other research promoting humor as a positive characteristic [21]. At the same time, it is essential when designing a chatbot to consider research showing that people do not communicate in the same way with a chatbot as with people [10, 30, 36].

Moreover, the results showed that sophisticated choices of words and well-constructed sentences were preferred characteristics in a chatbot. This not only supports previous findings [26], but also extends them with data showing that specific and long sentences are desired. In addition, correct language usage was rated higher than everyday language that included abbreviations and incorrect punctuation. One explanation behind this may be that people, while interacting with a chatbot, expect communication to be similar to that of what they experience in a physical and formal context. Another could be context dependence, and valid when users are using chatbots in service settings (as opposed to for entertainment).

Further, a chatbot's dynamic approach in conversations may be more complex than previously thought. Specifically, having a dynamic approach does not only entail asking follow-up questions, but also being capable of staying in context and asking said questions by making use of information previously given by users [21, 37, 47]. The results demonstrated that follow-up questions may not be desired when asking for general information (e.g., weather information), but are preferred in decision-making processes (e.g., purchase choices). This aspect is essential, and must be carefully considered in the design process of a chatbot as it may have serious practical implications, for instance for users who may ask medical questions and receive recommendations based on their input [4, 18].

Past research has identified that expressing appropriate emotions when it is relevant in a given context may increase the perceived humanness in a chatbot in terms of competence and warmth [10]. The findings of this study support previous research in that users appreciate when chatbots express emotions when facing context they do not understand and when a chatbot follows up with a question or a statement to move a conversation forward. It was, however, not desired that a chatbot communicate an apology via a longer response.

The results of this study have shown, in line with previous studies [26, 30], that there exists a wide range of preferences of how users prefer to interact with a chatbot. To accommodate individual users, one way forward could be to create chatbots that rely on users' personality traits in concert with their specific preferences [30].

This study has certain limitations that need to be considered in future work. The survey was conducted with participants who were relatively familiar with the Internet, and specifically chatbots. Future work should aim to examine those who do not have much experience with technology in order to bridge a possible digital divide. The questions in the survey investigated specific situations and contexts. While these showed satisfactory internal reliability, they could have been randomized to minimize question order bias and improve overall data quality.

6 CONCLUSIONS

As the use of chatbots continues to spread, this will create new challenges for designers. While it has been common to design graphical user interfaces, a shift may be underway with greater focus on designing voice-activated interfaces rather than graphical ones. It will therefore become increasingly important for designers to better understand a user's intentions, needs, and goals to help them access information and perform actions through conversations, whether written or verbal, instead of graphical user interfaces. This transition may require a different set of methods to measure, analyze, design and deliver a positive user experience.

Another challenge is connected to designing interactions that help users create a correct mapping of a chatbot's intelligence and competence. This could be vital in order to avoid misunderstanding and disappointment when a chatbot does not live up to expectations. Chatbots provide the possibility for users to interact with technology on a personal level, potentially creating various kinds of expectations and different ways to fulfill their needs. These expectations and different approaches may need to be anticipated and prepared for in the design process of a chatbot, in order to accommodate a wider group of users. The results of this study and the proposed guideline contribute with theoretical pointers that could guide the future design of chatbots and which may have practical implications for the increasing adoption of AI in different parts of society.

7 REFERENCES

[1] C. Adam and B. Gaudou. 2016. BDI agents in social simulations: A survey. *The Knowledge Engineering Review* 31, 3 (2016), 207–238.

[2] N. Akma, M. Hafiz, A. Zainal, M. Fairuz, and Z. Adnan. 2018. Review of chatbots design techniques. *International Journal of Computer Applications* 181, 8 (2018), 7–10.

[3] T. Bickmore and J. Cassell. 2005. *Social dialogue with embodied conversational agents.* Springer, Dordrecht, NL, 23–54.

[4] T. Bickmore, H. Trinh, S. Olafsson, T. K. O'Leary, R. Asadi, N. M. Rickles, and R. Cruz. 2018. Patient and consumer safety risks when using conversational assistants for medical information: An observational study of Siri, Alexa, and Google Assistant. *Journal of Medical Internet Research* 20, 9 (2018), e11510.

[5] A. Braga and R. Logan. 2017. The emperor of strong AI has no clothes: Limits to artificial intelligence. *Information* 8, 4 (2017), 156–177.

[6] C. Chakrabarti and G. F. Luger. 2015. Artificial conversations for customer service chatter bots: Architecture, algorithms and evaluation metrics. *Expert Systems with Applications* 42, 2015 (2015), 6878–6897.

RIGHTSLINK

[7] Swedish Research Council. 2017. *Good research practice*. Report. Swedish Research Council.

[8] L. J. Cronbach. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 3 (1951), 297–334.

[9] R. Dale. 2016. The return of the chatbots. *Natural Language Engineering* 22, 5 (2016), 811–817.

[10] V. Demeure, R. Niewiadomski, and C. Pelachaud. 2011. How is believability of a virtual agent related to warmth, competence, personification, and embodiment? *Presence* 20, 5 (2011), 431–448.

[11] A. Dix. 2010. Human-computer interaction: A stable discipline, a nascent science, and the growth of the long tail. *Interacting with Computers* 22, 1 (2010), 13–27.

[12] P. Dourish. 2004. What we talk about when we talk about context. *Personal Ubiquitous Computing* 8, 1 (2004), 19–30.

[13] D. Duijst. 2017. *Can we improve the user experience of chatbots with personalisation?* Thesis.

[14] A. Følstad and P. B. Brandtzaeg. 2017. Chatbots and the new world of HCI. *Interactions* 24, 4 (2017), 38–42.

[15] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena. 2018. AI4People - An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines* 28, 4 (2018), 689–707.

[16] N. Foulquier, P. Redou, C. Le Gal, B. Rouvière, J-O. Pers, and A. Saraux. 2018. Pathogenesis-based treatments in primary Sjogren's syndrome using artificial intelligence and advanced machine learning techniques: A systematic literature review. *Human Vaccines and Immunotherapeutics* 14, 11 (2018), 2553–2558.

[17] M. Gams, I. Y. Gu, A. Härmä, A. Muñoz, and V. Tam. 2019. Artificial intelligence and ambient intelligence. *Journal of Ambient Intelligence and Smart Environments* 11, 1 (2019), 71–86.

[18] T. Hagendorff and K. Wezel. 2019. 15 challenges for AI: or what AI (currently) can't do. *AI and Society* 3, 2019 (2019), 1–11.

[19] J. Hecht. 2018. Meeting people's expectations. *Nature Outlook - Digital Revolution* 563, 7733 (2018), 141–143.

[20] M. Huang and R. Rust. 2017. Artificial intelligence in service. *Journal of Service Research* 21, 2 (2017), 155–172.

[21] M. Jain, P. Kumar, R. Kota, and S. N. Patel. 2018. Evaluating and informing the design of chatbots. In *DIS 2018, Session 18: Interacting with Conversational Agents*. 895–906.

[22] M. H. Jarrahi. 2018. Artificial intelligence and the future of work: Human-AI symbiosos in organizational decision making. *Business Horizons* 61, 4 (2018), 577–586.

[23] L. C. Klopfenstein, S. Delpriori, S. Malatini, and A. Bogliolo. 2017. The rise of bots: A survey of conversational interfaces, patterns and paradigms. In *DIS 2017*. 555–565.

[24] C. Lallemand, G. Gronier, and V. Koenig. 2015. User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey. *Computers in Human Behavior* 43, 2015 (2015), 35–48.

[25] Q. V. Liao, M. Hussain, P. Chandar, M. Davis, Y. Khazaen, M. P. Crasso, D. Wang, M. Muller, N. S. Shami, and W. Geyer. 2018. All work and no play? Conversations with a question-and-answer chatbot in the wild. In *CHI 2018*. 1–13.

[26] C. L. Lortie and M. J. Guitton. 2011. Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS ONE* 6, 9 (2011), e25085.

[27] E. Luger and A. Sellen. 2016. "Like having a really bad PA": The gulf between user expectation and experience of conversational agents. In *CHI 2016*. 5286–5297.

[28] S. Makridakis. 2017. The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures* 90, 2017 (2017), 46–60.

[29] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon. 1955/2006. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine* 27, 4 (1955/2006), 12–14.

[30] D. McDuff and M. Czerwinski. 2018. Designing emotionally sentient agents. *Commun. ACM* 61, 12 (2018), 74–83.

[31] S. McKenney, N. Nieveen, and J. van den Akker. 2006. *Design research from a curriculum perspective*. Routledge, London, 67–90.

[32] M. L. McNeal and D. Newyear. 2013. Introducing chatbots in libraries. *Library Technology Reports* 49, 8 (2013), 5–10.

[33] G. Mone. 2016. The edge of the uncanny. *Commun. ACM* 59, 9 (2016), 17–19.

[34] M. Mori. 1970. Bukimi no tani [The uncanny valley]. *Energy* 7, 4 (1970), 33–35.

[35] M. Mori. 2012. The uncanny valley. *IEEE Robotics and Automation Magazine* 19, 2 (2012), 98–100.

[36] Y. Mou and K. Xu. 2017. The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior* 72, 2017 (2017), 432–440.

[37] M. Neururer, S. Schlögl, L. Brinkschulte, and A. Groth. 2018. Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction* 2, 60 (2018), 1–19.

[38] S. Noorunnisa, D. Jarvis, J. Jarvis, and M. Watson. 2019. Application of the GO-RITE BDI framework to human-autonomy teaming: A case study. *Journal of Computing and Information Technology* 27, 1 (2019), 13–24.

[39] E. Norling. 2004. Folk psychology for human modelling: Extending the BDI paradigm.. In *AAMAS '04 Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*. 202–209.

[40] E. Paikari and A. van der Hoek. 2018. A framework for understanding chatbots and their future. In *Proceedings of 11th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE'18)*. 13–16.

[41] K. Panetta. 2017. Gartner top strategic predictions for 2018 and beyond. (2017). Retrieved September 5, 2019 from https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/

[42] J. Pereira and Ó. Díaz. 2019. Using health chatbots for behavior change: A mapping study. *Journal of Medical Systems* 43, 5 (2019), 1–13.

[43] L. Piccolo, M. Mensio, and H. Alani. 2018. Chasing the chatbots: Directions for interaction and design research. In *CONVERSATIONS 2018, 5th International Conference on Internet Science*. 1–12.

[44] M. Portela and C. Granell-Canut. 2017. A new friend in our smartphone? Observing interactions with chatbots in the search of emotional engagement. In *Interacción '17*. 1–7.

[45] R. Rosales, M. Castañón-Puga, L. Lara-Rosano, D. R. Evans, N. Osuna-Millan, and M. V. Flores-Ortiz. 2017. Modelling the interruption on HCI using BDI agents with the fuzzy perceptions approach: An interactive museum case study in Mexico. *Applied Sciences* 7, 8 (2017), 1–18.

[46] S. Russell and P. Norvig. 2010. *Artificial intelligence: A modern approach*. Pearson, Upper Saddle River, NJ.

[47] R. Schuetzler, M. Grimes, J. S. Giboney, and J. Buckman. 2014. Facilitating natural conversational agent interactions: Lessons from a deception experiment. In *35th International Conference on Information Systems*. 1–16.

[48] M. Skjuve, I. M. Haugstveit, A. Følstad, and P. B. Brandtzaeg. 2019. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Human Technology* 15, 1 (2019), 30–54.

[49] M. Strait, L. Vujovic, V. Floerke, M. Scheutz, and H. Urry. 2015. Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. In *33rd Annual ACM Conference on Human Factors in Computing Systems*. 3593–3602.

[50] J. Torresen. 2018. A review of future and ethical perspectives of robotics and AI. *Frontiers in Robotics and AI* 4, 1 (2018), 1–10.

[51] A. Turing. 1950. Computing machinery and intelligence. *Mind* 59, 236 (1950), 433–460.

[52] J. J. H. van den Akker. 1999. *Principles and methods of development research*. Kluwer Academic Publishers, Dordrecht.

[53] A. Vinciarelli, A. Esposito, E. André, F. Bonin, M. Chetouani, J. F. Cohn, M. Cristani, F. Fuhrmann, E. Gilmartin, Z. Hammal, D. Heylen, R. Kaiser, M. Koutsombogera, A. Potamianos, S. Renals, G. Riccardi, and A. A. Salah. 2015. Open challenges in modelling, analysis and synthesis of human behaviour in human-human and human-machine interactions. *Cognitive Computation* 7, 4 (2015), 397–413.

[54] J. B. Walther. 2007. Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior* 23, 5 (2007), 2538–2557.

[55] K. Warwick and H. Shah. 2016. Passing the Turing Test does not mean the end of humanity. *Cognitive Computation* 8, 3 (2016), 409–419.

[56] D. Westerman, A. C. Cross, and P. G. Lindmark. 2018. I believe in a thing called bot: Perceptions of the humanness of "chatbots". *Communication Studies* 70, 3 (2018), 1–18.

[57] Y. Yang, X. Ma, and P. Fung. 2017. Perceived emotional intelligence in virtual agents. In *CHI '17*. 2255–2262.

[58] J. Zamora. 2017. I'm sorry, Dave, I'm afraid I can't do that: Chatbot perception and expectations. In *HAI 2017*. 253–260.

RIGHTSLINK

APPENDIX A

- My age is ...
- I am a ...
  - Woman
  - Man
  - Other option
  - Do not wish to disclose
- Number of hours I spend on the internet every day:
  - 0-1
  - 2-5
  - 6-10
  - 11-20
- Do you use a smartphone?
  - Yes
  - No
- How much experience do you have with interacting with chatbots?
  The definition of a chatbot in this survey is a virtual personal assistant that you interact with through written or spoken language.
  - None
  - A few times a year
  - A few times a month
  - A few times a week
  - Every day
- I have used a chatbot in the following situations:
  You can check more than one option.
  - Searching for information
  - Booking meetings
  - Asking for the weather
  - Playing music
  - To have someone to talk to
  - To get inspiration when shopping online
  - Customer service
  - Surveys/evaluations
  - Job related situations
  - None of the above
  - If other, please specify
- If you need help when visiting a website, do you ask for help in the chat (if there is one)?
  - Never
  - Sometimes
  - All the time

RIGHTSLINK

APPENDIX B

Table 1: Number of responses and corresponding frequency for factors related to perceived humanness in chatbots.

Factor	Question	Responses	Frequency
F1	**1. I prefer the following greeting by a chatbot:**		
	(a) By saying hello.	9	13%
	(b) By saying hello, followed by my name.	7	10.1%
	(c) By saying hello, followed by my name and asking me what I want help with.	**26**	**37.7%**
	(d) It should be my initiative to start a conversation.	**27**	**39.1%**
F5	**2. I prefer the following response when a chatbot does not understand me:**		
	(a) "I am sorry, I do not understand."	9	13%
	(b) "I am sorry, I do not understand. Maybe you can help me by asking your question in a different way?"	**28**	**40.6%**
	(c) "I am sorry, I do not understand. Do you want to talk with a human instead?"	**31**	**44.9%**
	(d) "I do not know what you mean, try again!"	1	1.3%
	(e) "I have searched the internet for "[your previous question/statement]".	0	0%
F1	**3. You have asked a chatbot to help you look for destinations for your next business trip. Which update message do you prefer?**		
	(a) "Back soon."	2	2.9%
	(b) "Searching for destinations..."	**46**	**66.7%**
	(c) "Going through all the exciting options."	7	10.1%
	(d) "I don't want an update, only the result when the chatbot has finished the search."	6	8.7%
	(e) "Oh, a trip! *clapping hands* Can I come with? Think about it while I look up the best options for your destination."	7	10.1%
	(f) "Drifting away, dreaming about the perfect destination, back soon..."	1	1.4%
F3	**4. You ask a chatbot when the next bus to a train station leaves. Which response do you prefer?**		
	(a) "12:15"	5	7.2%
	(b) "Bus 11 towards *destination* leaves at 12:15."	**30**	**43.5%**
	(c) "Bus 11 leaves at 12:15, Bus 24 leaves at 12:18 and Bus 11 leaves at 12:25."	**27**	**39.1%**
	(d) "There are five different buses leaving to the train station from your bus stop, number 11, number 24, number 4, number 13 and number 121. Bus 11 leaves at 12:15, bus 24 leaves at 12:18, bus 4 leaves at 12:26, bus 13 leaves at 12:30 and bus 121 leaves at 12:31."	7	10.1%
F4	**5. You have asked a chatbot for the weather tomorrow. Which response do you prefer?**		
	(a) "In what city?"	2	2.9%
	(b) "15 degrees celsius."	0	0%
	(c) "15 degrees celsius and sunny with some clouds."	10	14.5%
	(d) "In Gothenburg it will be 15 degrees celsius and sunny with some clouds tomorrow."	**57**	**82.6%**
F2	**6. You are interacting with a chatbot that is connected to your bank. Which opening message do you prefer?**		
	(a) "Yo, where's the money?"	2	2.9%
	(b) "Hello, how can I help you today?"	**29**	**42%**
	(c) "Hi! Are you looking to save or spend today?"	5	7.2%
	(d) "Good morning. I am your banking assistant bot. How can I be of service today?"	**33**	**47.8%**
F2	**7. You are interacting with a bot that helps you play music. Which opening message do you prefer?**		
	(a) "Please state genre, artist or song title in order to receive appropriate choices."	5	7.2%
	(b) "Tell me what you want to listen to, and I can help you find it!"	**37**	**53.6%**
	(c) "Hi! Good to see you again, which genre do you feel like listening to today?"	**21**	**30.4%**
	(d) "Groovy, you're back! I have some sweet tunes for you, you ready to get down?"	6	8.7%

RIGHTSLINK

Continued from previous page

Factor	Question	Responses	Frequency
F2	**8. You are interacting for the first time with a chatbot at work to book a meeting. Which opening message do you prefer?**		
	(a) "Hi, I am your meeting buddy! What is on your mind? :)"	3	4.3%
	(b) "Hello! Nice to see you here, let me know how I can help you."	11	15.9%
	(c) "Do you want to book a room or manage your calendar?"	**21**	**30.4%**
	(d) "Welcome, would you like to see available times in your calendar?"	13	18.8%
	(e) "Welcome, I am a bot that can help you to manage your calendar and book a meeting room. Do you want me to suggest appropriate times for a meeting?"	**21**	**30.4%**
F3	**9. You have booked a trip with a chatbot. Which response do you prefer?**		
	(a) "Trip booked leaving tomorrow 09:05 arriving 12:08 in Stockholm."	9	13%
	(b) "Your trip is booked. Leaving tomorrow at 09:05 and arriving 12:08 in Stockholm."	**22**	**31.9%**
	(c) "Your trip is booked. Your train departs from the station tomorrow at 09:05 and will be arriving at 12:08 in Stockholm."	**38**	**55.1%**
F5	**10. You are frustrated that a chatbot can not understand what you want and have expressed the frustration to the chatbot. Which response do you prefer?**		
	(a) "I don't understand. Please try again."	14	20.3%
	(b) "I am sorry, it seems I am not following, could you please rephrase and try again?"	**26**	**37.7%**
	(c) "I am so sorry, I really want to help you but I just can not understand. Please rephrase or try simple keywords instead."	**24**	**34.8%**
	(d) "I am terribly sorry, this is not how I wanted it to end up either..! I am doing my best to "help you, please bare with me, I will do everything in my power to make it up to you."	5	7.2%
F4	**11. You ask a chatbot to help you choose between two hair trimmers. Which response do you prefer?**		
	(a) "Buy the second option."	3	4.3%
	(b) "Buy the second option, it is rated highest by customer reviews."	**31**	**44.9%**
	(c) "Do you need a trimmer for body, beard or for hair?"	**32**	**46.4%**
	(d) "You bought a trimmer for body five months ago, is this one for beard or hair?"	3	4.3%

Footnotes: Each question was related to one or more of the five factors in this study, namely: (F1) small talk; (F2) way of conduct in tasks; (F3) sophisticated choices of words and well-constructed sentences; (F4) dynamic approach; and (F5) expressing feelings based on context.

RIGHTSLINK