

Measuring user experience in chatbots: An approach to interpersonal communication competence

Marita Skjuve¹ and Petter Bae Brandzaeg¹

¹ SINTEF, OSLO, Norway
marita.skjuve@sintef.no

The increasing usage of chatbots is fundamentally changing the way people interact with new technology. Instead of clicking buttons to functionally navigate on a web page, people can access content and services by the use of natural language in interaction with an artificial agent (e.g., chatbot). This change toward human–chatbot interaction is typically manifested through a social and natural conversational style. This shift of how to interact with data and services has major repercussions for how to explore and measure conversational user experience with chatbots. Hence, in this paper we suggest the importance of measuring the social aspects in human–chatbot interaction through a focus on interpersonal communicational competence (ICC). We build our suggested framework on previous knowledge about communicational competence in successful human–human interaction. Based on this we have developed factors that might be of importance for chatbots’ ICC and how these can be measured.

Keywords: Chatbot, Conversational interface, Natural language, Human–chatbot interaction, User experience, Interpersonal communicational competences

1 Introduction

The ability to demonstrate interpersonal communicational competence (ICC) is critical for functional human–human interaction and relationships. Technology, up to now, has not been required to demonstrate this ability. However, the emergence of conversational interfaces such as chatbots demands new methods in the field of human–computer interaction (HCI) to measure conversational user experiences when interacting with chatbots. Chatbots are understood as machine agents serving as conversational user interfaces to data and services online [1]. The term “*chatbot*” is just one of many names describing this software. Other names include, but are not limited to, conversational interface, conversational agent, and intelligent personal assistant.

Chatbots may utilize artificial intelligence, machine learning, and natural language processing, which makes them capable of communicating with the user the same way humans communicate with each other—through natural language. This results in users no longer having to click their way through a web page. They can simply talk to the chatbot and it will provide them with the information needed.

It is suggested that chatbots will involve a fundamental break with past user interfaces and a change in user motivations and interaction patterns online [2]. This change in future interactions and user interfaces requires a radical rethinking about how to measure the user experiences of chatbots. Chatbots can be regarded as social technologies, and HCIs therefore need to treat them as such rather than just functional technologies. While there has been an awareness of the importance of supporting enjoyable social interaction when designing communication systems [3], this topic is still not fully explored in the user experience literature [4] and certainly not in the context of chatbots.

Chatbots are deemed to be a promising conversational user interface to data and services for several reasons. They are efficient, cost-effective, available, and user-friendly, and are now being “employed” in a variety of different positions. That is, they are working as customer service agents, as health advisors, as therapists, and as teachers. Some can even become your new internet best friend forever (BFF) or significant other. Chatbot developers must therefore be open-minded when thinking about relational development and social trust in human–chatbot interaction.

Chatbots are increasingly stepping into roles that previously belonged to humans. We know from previous research that humans tend to interact with artificial agents much in the same way they interact with humans [5]. This sets the tone for the ability these chatbots need in order to provide the user with a pleasurable interaction and user experience. But how can such user experience be assessed?

User experience is a complex term and is understood and measured in several ways [6], often because it is regarded as a subjective experience [4]. When measuring user experience in the human–chatbot interaction, we cannot rely only on prior usability scales such as a system usability scale (e.g., presented in [7]) simply because the user is not interacting with the technology the same way as they did before. Other types of scales are therefore frequently used, such as variations of the Godspeed questionnaire [8], which measures human likeness, and Social Presence scales, which measure aspects such as how warm or cold, and personal or impersonal a given media is [9].

These scales, presented above, are valuable ways of measuring different user experience aspects of the chatbot and ICC, as presented herein, may affect variables covered in these scales. For example, it is likely that a chatbot that has high ICC is perceived as more humanlike or to have a higher degree of social presence because it communicates in a warmer and more personal way.

New and broader ways of measuring user experience in the human–chatbot interaction are therefore necessary to cover the more conversational nature of chatbots. The aim should be to cover the experience of the communication and exchange process of both information and feelings by means of verbal or non-verbal messages, between the chatbot and the user.

The objective of this paper is therefore to (1) identify factors in previous literature that have been demonstrated to be important for successful communication in human–human interaction, and (2) present an initial scale on how ICC, understood as interpersonal competence, can be measured in chatbots. This work may help us to investigate and measure user experience factors that affect how users perceive human–chatbot interaction in the future.

2 What is the recipe for a successful conversation?

Spitzberg and Hecht [10] argue that the level of communication competence sets the tone for how an interaction is perceived. According to the authors, communication between two or more individuals needs to be appropriate to the context it is carried out in, and state that “competent communication is a process through which interpersonal impressions are shaped and satisfactory outcomes are derived from an interaction” (p. 576).

Spitzberg and Hecht [10] go on and present the following four important constructs in interpersonal relationships: (1) communication motivation, (2) knowledge, (3) skills, and (4) outcomes. These factors are further argued to be a universal part of creating an impression of having high communicational competence.

Rubin and Martin [11] echo this and state that “interpersonal communication competence (ICC) is an impression or judgment formed about a person’s ability to manage interpersonal relationships in communication settings.” While Spitzberg and Hecht [10] identify four competences, Rubin and Martin [11] explain that research over the last 20 years has tended to agree upon a total of 10 ICC skills (see Table 1). These competences overlap with the four presented by Spitzberg and Hecht [10], and have been used to test ICC in a variety of settings—ranging from attachment and social support [12] to acceptance of robots [13].

While previous research has tended to focus on measuring ICC in human–human interaction, some studies have applied this in user experience tests with artificial agents, such as robots. Maartje De Graaf, et al. [14], for instance, showed that participants who found their robot to lack sociability were more likely to stop using it.

After reviewing the list of competences from Rubin and Martin [11] and the previous chatbot literature, we argue that the following ten competences are important in a human–chatbot interaction (presented in Table 1).

Table 1. Communicational competences important for the user experience.

Competency	Explanation	Relevance
Self-disclosure	The chatbot should be able to share personal thoughts or experiences.	Transcripts from human–chatbot interactions indicate that the user asks the chatbot questions about itself [15]. The chatbot should thus be able to share information when appropriate.
Empathy	The chatbot should be able to demonstrate that it understands and/or feels with the user when appropriate.	Fitzpatrick, et al. [16] found that the user valued empathic responses in the chatbot, and Kim, et al. [17] showed that teenagers want a chatbot to “empathize with me.”
Social relaxation	The chatbot should feel comfortable and secure during the interaction, and not be anxious.	Marrinan [18] found that the participants liked that the chatbot seemed calm and relaxed, but also appreciated its “expert” and more professional demeanor.
Interaction management	The chatbot should be able to demonstrate turn-taking and discuss and develop different topics.	Luger and Sellen [19] found that users often complained about the chatbot’s lack of ability to discuss or understand follow-up questions to a topic.

Assertiveness	The chatbot should be able to stand up for itself and its own right, but at the same time be accustomed to, and not violate, the user's right.	Curry and Rieser [20] point out that chatbots should be able to use mitigation strategies to avoid facilitating harassment.
Altercentrism	The chatbot should make the user feel that it is interested in what he/she has to say, ask appropriate questions, be polite, and display appropriate emotional expressions and content.	Kim et al. [17] found that teenagers valued that the chatbot does not get tired of listening to them. They also stated that they want a chatbot that provides appropriate responses.
Expressiveness	The chatbot should be able to express its feelings either verbally (e.g., laughter) or non-verbally (emojis).	Fitzpatrick et al. [16] used emojis in their bot to facilitate the chatbot's emotional expressions.
Immediacy	The chatbot should be available and open for communication.	Fitzpatrick et al. [16] found that users liked that the chatbot automatically checked up on them, and Kim et al. [17] showed how "active listening" capabilities were deemed important.
Supportiveness	The chatbot should not judge the user and should make the user feel to be an equal.	Kim et al. [17] found that teenagers valued that the chatbot does not judge or get offended.
Environmental control	The chatbot should be able to accomplish its goals and objectives.	Luger and Sellen [19] argue that chatbots should be clear about explaining what it can and cannot do in order to set the right expectations.

Not all aspects related to the social competences listed in this table may be relevant for all types of chatbots. Different types of chatbots can be identified, based on the way in which they interact with the user, the content, and the context of the application [21]. We should therefore remember that chatbots come in different forms, as well with a variety of social and informational purposes. Some chatbots are in addition used over time, while others are relevant only for one immediate short interaction, such as information about today's weather. Hence, the type of dimension used and measured for different contexts should be related to the various social purposes and general context of the chatbot.

3 How can we measure the chatbot's ICC?

Rubin and Martin [11] present an ICC Scale. We have revised this scale to make it appropriate for measuring the chatbot's ICC. The questions were formulated slightly differently and two items were left out due to a lack of relevance (item 9: "I feel insecure in groups of strangers" and item 26: "I try to look others in the eye when speaking with them").

Table 2. Initial scale measuring each ICC.

Competency	Question
Self-disclosure	The chatbot gave me a sense of who it was
	The chatbot revealed what it was thinking
	The chatbot shared its feelings with me
Empathy	The chatbot seemed to know how I was feeling
	The chatbot seemed to understand me

	The chatbot put itself in my shoes
Social relaxation	The chatbot seemed to be comfortable talking with me The chatbot seemed relaxed and secure when talking with me
Interaction management	The chatbot took charge of the conversation The chatbot negotiated the topics we were discussing The chatbot seemed to pick up my non-verbal cues such as emojis The chatbot managed to shift smoothly between topics
Assertiveness	The chatbot confronted me when I was rude The chatbot had trouble standing up for itself when I was rude to it
Altercentrism	The chatbot managed to stay focused during the conversation The conversation felt one-sided The chatbot let me know that it understood what I was saying
Expressiveness	The chatbot let me know when it was happy or sad The chatbot didn't have difficulty finding words to express itself The chatbot was able to express itself verbally
Supportiveness	I would describe the chatbot as a "warm" communication partner The chatbot did not judge me The chatbot communicated with me as though we were equals
Immediacy	The chatbot made me feel like it cared about me The chatbot made me feel close to it
Environmental control	The chatbot seemed to accomplish its communication goals The chatbot managed to persuade me to its position The chatbot didn't have trouble convincing me to do what it wanted me to do

Note: Each competence is measured on a Likert scale, ranging thus: 1 = Not true at all, 2 = Mostly not true, 3 = neither true or not true, 4 = Mostly true, 5 = Very true

4 Conclusion and future directions

This paper is an initial attempt to present a new way of measuring user experience in terms of ICC—the interpersonal competence in chatbots. The 10 ICC skills and the ICC scale presented by Rubin and Martin [11] to measure interpersonal competence in human–human interaction were described and tailored to the context of human–chatbot interaction. During the process of tailoring the scale, some observations were made:

First, items measuring “assertiveness” may need to be specified and include references to cases such as harassment. It might be difficult for the user to understand “why the chatbot should stand up for itself.” While this may not seem relevant in its current form, we argue that it is relevant in the context of harassment. We have several examples of how users have harassed chatbots in the past [20] and Curry and Rieser [20] argue that it is important that the chatbot mitigate such tendencies to avoid facilitating inappropriate behaviors. Thus, this ICC might be viewed as a necessary competence for regulating the behavior of the user, in cases where that is appropriate.

Second, “social relaxation,” as understood by Rubin and Martin [11], might be more about the stress one feels, than the behavior one expresses. The understanding of this factor in a chatbot context may thus need some alteration where the emphasis is more on the ability to express confidence more so than to “feel” relaxed.

Third, the factor “immediacy” is seemingly overlapping with the factor empathy. One suggestion is to focus more on the ability to demonstrate availability through push messages and active listening.

Fourth, because the conversation with the chatbot is computer-mediated, some items addressing this should be added to the scale. This is also in line with Spitzberg [22]. Under “interaction management,” for example, items such as (1) the length of the responses was appropriate, (2) the number of responses was appropriate, and (3) the response time was appropriate, would arguably be relevant for the chatbot’s ICC.

4.1 Future directions

We encourage future work to continue to build on this potential framework for measuring a chatbot’s ICC. This includes, among other things to:

1. review the suggested measure in this study against other scales and measures (e.g., social presence, or the Godspeed questionnaire) that are increasingly being used to evaluate chatbots in the social robot literature.
2. validate the scale and the factors. Here it will be important to ask whether the scale looks sound, whether it asks about the sorts of thing that we think of as being related to social competence in relation to chatbots (face validity). The scale needs to be tested in terms of: (1) the property of having appropriate relationships with other variables (construct validity) (2) and to what degree the items that compose the scale are related to one another (internal consistency).
3. test the scale among several observers that apply the scale independently, also for a range of various chatbots that need to apply interpersonal competence.
4. Explore other factors and develop guidelines for the contextual importance of each factor. Because the significance of the different factors will vary according to the context and the role the chatbot performs, not all the dimensions we have suggested in this paper will be relevant for all types of chatbots. Self-disclosure might, for instance, be more important in chatbots that function as a friend, than a customer service chatbot. We also need to investigate how long human–chatbot interaction is necessary in order to judge the chatbot’s ICC.

References

1. Dale, R. The return of the chatbots. *Natural Language Engineering*, 22, 5 (2016), 811-817.
2. Brandtzaeg, P. B. and Følstad, A. Chatbots: changing user needs and motivations. *Interactions*, 25, 5 (2018), 38-43.
3. Monk, A. *User-centred design*. Springer, City, 2000.
4. Brandtzæg, P. B., Følstad, A. and Heim, J. *Enjoyment: lessons from Karasek*. Springer, City, 2018.
5. Nass, C., Steuer, J. and Tauber, E. R. *Computers are social actors*. ACM, City, 1994.

6. Bargas-Avila, J. A. and Hornbæk, K. *Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience*. ACM, City, 2011.
7. Brooke, J. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189, 194 (1996), 4-7.
8. Ho, C.-C. and MacDorman, K. F. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26, 6 (2010), 1508-1518.
9. Biocca, F., Harms, C. and Burgoon, J. K. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, 12, 5 (2003), 456-480.
10. Spitzberg, B. H. and Hecht, M. L. A component model of relational competence. *Human Communication Research*, 10, 4 (1984), 575-599.
11. Rubin, R. B. and Martin, M. M. Development of a measure of interpersonal communication competence. *Communication Research Reports*, 11, 1 (1994), 33-44.
12. Anders, S. L. and Tucker, J. S. Adult attachment style, interpersonal communication competence, and social support. *Personal Relationships*, 7, 4 (2000), 379-389.
13. De Graaf, M. M. and Allouch, S. B. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61, 12 (2013), 1476-1486.
14. De Graaf, M., Ben Allouch, S. and Van Dijk, J. *Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study*. ACM, City, 2017.
15. Thies, I. M., Menon, N., Magapu, S., Subramony, M. and O'neill, J. *How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians*. Springer, City, 2017.
16. Fitzpatrick, K. K., Darcy, A. and Vierhile, M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, 4, 2 (2017).
17. Kim, J., Kim, Y., Kim, B., Yun, S., Kim, M. and Lee, J. *Can a Machine Tend to Teenagers' Emotional Needs?: A Study with Conversational Agents*. ACM, City, 2018.
18. Marrinan, F. *Qualitative investigations into a virtual CBT therapist: relational features, client experiences and implications for counselling psychology practice*. University of Surrey, 018.
19. Luger, E. and Sellen, A. *Like having a really bad PA: the gulf between user expectation and experience of conversational agents*. ACM, City, 2016.
20. Curry, A. C. and Rieser, V. *# MeToo Alexa: How Conversational Systems Respond to Sexual Harassment*. City, 2018.
21. Di Gaetano, S. and Diliberto, P. *Chatbots and Conversational Interfaces: Three Domains of Use* (2018).

22. Spitzberg, B. H. Preliminary development of a model and measure of computer-mediated communication (CMC) competence. *Journal of Computer-Mediated Communication*, 11, 2 (2006), 629-666.